

N° 842

SÉNAT

SESSION EXTRAORDINAIRE DE 2024-2025

Enregistré à la Présidence du Sénat le 9 juillet 2025

RAPPORT D'INFORMATION

FAIT

au nom de la commission de la culture, de l'éducation, de la communication et du sport (1) par la mission d'information sur l'intelligence artificielle (IA) et la création,

Par Mmes Agnès EVREN, Laure DARCOS et M. Pierre OUZOULIAS,

Sénatrices et Sénateur

(1) Cette commission est composée de : M. Laurent Lafon, président ; MM. Jérémy Bacchi, Max Brisson, Yan Chantrel, Mme Laure Darcos, MM. Bernard Fialaire, Jacques Gersperrin, Martin Lévrier, Mmes Monique de Marco, Marie-Pierre Monier, M. Michel Savin, vice-présidents ; Mmes Colombe Brossel, Else Joseph, M. Pierre-Antoine Levi, Mme Anne Ventalon, secrétaires ; Mmes Marie-Jeanne Bellamy, Catherine Belrhiti, Annick Billon, Alexandra Borchio Fontimp, M. Christian Bruyen, Mmes Samantha Cazebonne, Mireille Conte Jaubert, Evelyne Corbière Naminzo, Karine Daniel, Sabine Drexler, M. Aymeric Durox, Mmes Agnès Evren, Laurence Garnier, Béatrice Gosselin, MM. Jean Hingray, Patrick Kanner, Claude Kern, Mikaele Kulimoetoke, Mme Sonia de La Provôté, MM. Ahmed Laouedj, Michel Laugier, Jean-Jacques Lozach, Mmes Pauline Martin, Catherine Morin-Desailly, M. Georges Naturel, Mme Mathilde Ollivier, MM. Pierre Ouzoulias, Jean-Gérard Paumier, Stéphane Piednoir, Mme Sylvie Robert, MM. David Ros, Pierre-Jean Verzellen, Cédric Vial, Adel Ziane.

SOMMAIRE

	<u>Pages</u>
LISTE DES RECOMMANDATIONS DE LA MISSION D'INFORMATION	5
AVANT-PROPOS	7
I. PETIT PRÉCIS HISTORIQUE ET TECHNIQUE DE L'INTELLIGENCE ARTIFICIELLE	11
A. DES AUTOMATES ANTHROPOMORPHES DE L'ANTIQUITÉ À L'AGENT CONVERSATIONNEL CHATGPT : L'AVENTURE DE L'INTELLIGENCE ARTIFICIELLE	11
1. <i>Aux origines imaginaires et rationnelles de l'intelligence artificielle</i>	<i>11</i>
2. <i>Naissance et développement de l'intelligence artificielle : une histoire non linéaire</i>	<i>12</i>
B. GLOSSAIRE DE L'INTELLIGENCE ARTIFICIELLE.....	16
1. <i>L'intelligence artificielle et ses sous-domaines</i>	<i>17</i>
2. <i>Modèle d'intelligence artificielle vs système d'intelligence artificielle</i>	<i>21</i>
C. LA CONCEPTION ET LE FONCTIONNEMENT D'UN MODÈLE D'INTELLIGENCE ARTIFICIELLE GÉNÉRATIVE.....	22
1. <i>Une diversité de modèles et d'acteurs</i>	<i>22</i>
2. <i>Les différentes étapes de la conception d'un modèle d'intelligence artificielle générative.....</i>	<i>24</i>
II. LA CRÉATION ARTISTIQUE FACE À LA VAGUE DE L'INTELLIGENCE ARTIFICIELLE : NOUVEL HORIZON ARTISTIQUE OU PARASITISME CRÉATIF ?	29
A. COMMENT L'INTELLIGENCE ARTIFICIELLE BOULEVERSE LE PROCESSUS DE CRÉATION.....	29
1. <i>Des premières expériences artistiques recourant à l'intelligence artificielle aux premières œuvres intégralement générées par celle-ci</i>	<i>29</i>
2. <i>Une échelle de gradation de la place de l'intelligence artificielle dans le processus créatif...30</i>	<i>30</i>
3. <i>Principaux cas d'usage de l'intelligence artificielle dans les industries culturelles et créatives</i>	<i>32</i>
B. LA CONCURRENCE D'UNE NOUVELLE FORME DE CRÉATION QUI POSE UN DÉFI ÉCONOMIQUE, SOCIAL ET CULTUREL.....	35
1. <i>Les contenus culturels sortants : des productions parasitaires</i>	<i>35</i>
2. <i>Une menace pour les métiers de la création</i>	<i>37</i>
3. <i>Un risque d'uniformisation et de déshumanisation de la création artistique.....</i>	<i>38</i>
III. UN ENCADREMENT JURIDIQUE EN CONSTRUCTION AU NIVEAU MONDIAL.....	41
A. LA DIRECTIVE DU 17 AVRIL 2019 SUR LES DROITS D'AUTEUR : UN SOCLE DÉSORMAIS DÉPASSÉ.....	42
1. <i>Les exceptions « text en data mining » (TDM).....</i>	<i>42</i>
2. <i>Une exception détournée ?</i>	<i>44</i>
3. <i>L'intervention des tribunaux</i>	<i>45</i>
B. UNE LÉGISLATION EUROPÉENNE QUI CHERCHE UNE DIRECTION.....	49
1. <i>La nécessité d'un cadre, la difficulté d'y parvenir</i>	<i>49</i>
2. <i>Les avancées et les incertitudes du RIA</i>	<i>51</i>

C. EN AVAL : QUEL STATUT JURIDIQUE POUR LES ŒUVRES GÉNÉRÉES PAR UNE IA ?	60
1. La définition classique du droit d'auteur	60
2. Le défi juridique posé par l'IA	61
IV. TRANSPARENCE ET RÉMUNÉRATION DANS L'INTÉRÊT DE TOUS.....	67
A. L'INSOUTENABLE GRATUITÉ DES DONNÉES	67
1. Selon la tech, des données gratuites par nature.....	67
2. La nature radicalement nouvelle des contenus culturels générés par l'IA.....	70
3. Une technologie de rupture qui justifie une évolution des règles.....	71
B. UNE SITUATION PRÉJUDICIABLE À TOUS	72
1. Des menaces pour les créateurs	73
2. Une double incertitude pour les entreprises technologiques	73
3. L'Europe, « idiot utile » des Américains ?.....	77
V. LES RECOMMANDATIONS DE LA MISSION D'INFORMATION POUR UN PARTAGE DE LA VALEUR ÉQUILIBRÉE ENTRE ACTEURS DE L'IA ET AYANTS DROIT CULTURELS, SOCLE SUR LEQUEL BÂTIR UNE TROISIÈME VOIE DE L'IA.....	79
A. L'ABSENCE PRÉOCCUPANTE D'UN MARCHÉ DES DONNÉES.....	79
B. HUIT GRANDS PRINCIPES À RESPECTER DANS LE CADRE DE L'ÉLABORATION CONCERTÉE D'UN MODÈLE DE RÉMUNÉRATION DES CONTENUS CULTURELS UTILISÉS PAR L'IA	82
1. Premier principe : le droit à rémunération des ayants droit est légitime et incontestable	82
2. Deuxième principe : la transparence sur les données utilisées doit être garantie.....	83
3. Troisième principe : la rémunération doit être fonction des flux de revenus générés par l'IA	84
4. Quatrième principe : la création de bases de données harmonisées, aux conditions d'utilisation clairement définies, est un préalable indispensable à l'existence d'un marché de la donnée.....	85
5. Cinquième principe : le passé doit être soldé	85
6. Sixième principe : un avantage comparatif doit être donné aux fournisseurs d'IA respectueux du cadre légal	86
7. Septième principe : la diversité culturelle et la créativité humaine doivent continuer à être encouragées	87
8. Huitième principe : les créations générées par l'IA doivent être étiquetées	87
C. DONNER SA CHANCE AU RIA ET GARANTIR SON EFFECTIVITÉ	88
1. Une mise en œuvre entravée.....	88
2. Une réponse graduée pour parvenir enfin à une rémunération appropriée des ayants droit culturels	88
EXAMEN EN COMMISSION.....	93
LISTE DES PERSONNES ENTENDUES ET DES CONTRIBUTIONS ÉCRITES	95
DÉPLACEMENT	101
TABLEAU DE MISE EN ŒUVRE ET DE SUIVI DES RECOMMANDATIONS	103

LISTE DES RECOMMANDATIONS DE LA MISSION D'INFORMATION

Assurer le respect des huit principes suivants, présentés sous forme de recommandations, dans la mise en place de relations équilibrées entre les ayants droit culturels et les fournisseurs d'IA :

Recommandation n° 1 : Réaffirmer et garantir le droit à rémunération des ayants droit culturels pour l'utilisation de leurs contenus par les fournisseurs d'IA.

Recommandation n° 2 : Garantir la transparence complète des données utilisées par les fournisseurs d'IA.

Recommandation n° 3 : Définir des modalités de rémunération qui soient fonction des flux de revenus générés par les fournisseurs et déployeurs d'IA.

Recommandation n° 4 : Inciter le secteur culturel et celui de la presse à constituer des bases de données larges et de qualité, facilement exploitables par les fournisseurs, assorties de conditions d'utilisation précisément définies.

Recommandation n° 5 : Parvenir à un règlement financier pour les usages passés des contenus culturels, afin de compenser les ayants droit culturels et sécuriser juridiquement les fournisseurs d'IA.

Recommandation n° 6 : Créer les conditions d'un réel avantage comparatif pour les fournisseurs d'IA vertueux qui sauront nouer les meilleurs accords avec les ayants droit culturels.

Recommandation n° 7 : Tirer profit des revenus générés par le marché de l'IA pour promouvoir la diversité de la création culturelle et le pluralisme de la presse.

Recommandation n° 8 : Travailler à la mise en place d'un système technique permettant d'identifier les contenus intégralement générés par l'IA.

Recommandation n° 9 : Garantir l'effectivité du droit d'auteur en suivant une réponse graduée :

- attente des conclusions, à l'automne prochain, de la concertation lancée par le ministère de la culture et le ministère de l'économie entre les fournisseurs d'IA et les ayants droit culturels ;
- en cas d'échec de cette concertation à trouver des solutions adaptées, dépôt d'une proposition de loi d'initiative sénatoriale visant à mettre en œuvre une présomption d'utilisation des contenus culturels par les fournisseurs d'IA ;
- en cas de nouvel échec, mise en place d'une taxation du chiffre d'affaires réalisé en France par les fournisseurs et déployeurs d'IA, afin de compenser le secteur culturel.

AVANT-PROPOS

La fascination de l'homme pour la création d'entités autonomes, dotées de caractéristiques humaines, irrigue toute l'histoire occidentale, des mythes de l'Antiquité à la cybernétique des années 1950, en passant par l'art de la Renaissance, le rationalisme philosophique du XVII^e siècle et la littérature de science-fiction du XIX^e siècle.

Si l'espoir de modéliser l'intelligence humaine se concrétise dans les années 1950 grâce aux progrès de l'informatique, l'optimisme des débuts de **l'intelligence artificielle (IA) - définie à la fois comme un domaine de recherche scientifique et une technologie** - s'est rapidement heurté aux capacités restreintes des premiers ordinateurs. S'en suit, au cours des décennies suivantes, **un parcours accidenté, fait de périodes de pics et de creux.**

Le développement, dans les années 2010, des techniques d'apprentissage profond, puis l'apparition, au début des années 2020, des **premières IA dites « génératives » marquent un changement de paradigme** : pour la première fois, **l'IA sort des sphères scientifique et technologique pour investir l'ensemble des pans de l'économie et de la société.**

Moins de trois ans après le lancement, le 30 novembre 2022, par la société OpenAI de son agent conversationnel **ChatGPT**, l'IA est passée de fantasme scientifique à réalité technologique pour des centaines de millions d'utilisateurs chaque jour sur toute la planète. En 1993, l'écrivain et mathématicien américain Vernor Vinge prédisait, à l'horizon de trente ans, la « *Singularité technologique* », soit le moment où l'intelligence de l'homme serait dépassée par celle de la machine. À quelques mois près, cette prophétie correspond à la mise sur le marché de ChatGPT.

En quelques années, **l'IA a réussi à pénétrer l'ensemble des strates de notre société** : l'organisation sociale, l'économie, la politique, mais aussi notre rapport aux autres. Elle occupe désormais **une place centrale dans le monde professionnel**, ouvrant la voie à des processus de réorganisation des entreprises et des administrations porteurs de promesses d'efficience et de croissance. Certains secteurs, comme la médecine ou la logistique, font d'ores et déjà un usage intensif de modèles d'IA spécialisés. **L'IA s'impose également dans l'espace privé**, avec l'utilisation en pleine expansion des grands modèles

de langage (LLM¹) tant pour structurer des informations que pour répondre à des demandes de leurs utilisateurs.

Cet essor fulgurant a pris de court nos modes de pensée et de régulation traditionnels. Le monde que l'IA promet ne s'insère en effet que difficilement dans nos systèmes économiques, juridiques et démocratiques.

Les concepteurs d'IA ont jusqu'à présent suivi un chemin déjà balisé par la précédente révolution numérique et largement inspiré de la doctrine libertarienne : agir vite de manière agressive, créer une situation de fait pour s'imposer, et renvoyer les conséquences à plus tard. **La conception et le fonctionnement des modèles d'IA reposent ainsi sur l'utilisation de quantités massives de contenus culturels, collectés sans que leurs détenteurs légitimes n'aient à aucun moment pu autoriser ou non leur exploitation, encore moins percevoir une rémunération appropriée.** Ce comportement peut être résumé par une formule de Grace Hopper, informaticienne américaine conceptrice des premiers compilateurs dans les années 1950 : « *Il est plus facile de demander pardon que de demander la permission*² ».

Si **le secteur culturel** est loin d'être le seul à être inquiété par la vague de l'IA, il est peut-être le plus **emblématique**. En effet, alors que les précédentes révolutions technologiques avaient largement pour conséquence de décharger l'homme de tâches pénibles ou d'exécution, **l'IA est désormais en capacité d'investir des domaines que l'on croyait jusqu'à présent réservés aux êtres humains.** Si une telle substitution est, dans certains domaines, un facteur précieux de connaissance et de progrès, elle peut, s'agissant du secteur culturel, constituer **une menace quasi existentielle dans la mesure où les productions générées par la machine entrent en concurrence directe avec les œuvres de l'esprit.** Cette évolution de nature anthropologique interroge profondément notre conception de l'humanité et réactive les peurs ancestrales sur le remplacement de l'Homme par la machine.

Les questions posées par l'IA sont d'autant plus complexes que ses conséquences ne se cantonnent pas aux seules sphères économique et sociale. L'IA est devenue **un enjeu géopolitique de souveraineté**, à tel point que les négociations douanières actuellement en cours entre les États-Unis et la Chine traitent pour partie autour de ce sujet : accès aux terres rares, aux puces, aux meilleurs talents...

Dans cette bataille pour la souveraineté numérique en matière d'IA, l'Europe est une nouvelle fois ramenée à ses faiblesses structurelles, déjà lourdement creusées par la révolution numérique des années 2000 au cours de laquelle notre continent a été plus spectateur qu'acteur. La maîtrise de notre destin en tant que continent souverain est directement liée à notre capacité à rattraper notre retard technologique en matière d'IA. Or cette prise de

¹ Large Language Model

² « *It's easier to ask forgiveness than it is to get permission.* »

conscience s'est accompagnée d'un mouvement presque de panique en faveur d'une absence de régulation, ou d'une régulation très limitée, présumée favorable au développement des acteurs de l'IA. Sur cette question, la France a souvent tenu un rôle plus ambigu que sa défense traditionnelle et inconditionnelle du droit d'auteur n'aurait pu le laisser présager.

C'est dans ce contexte disruptif que **la commission de la culture du Sénat a souhaité sortir du débat manichéen** qui oppose souvent, dans l'espace public, les défenseurs de la création aux thuriféraires de la technologie. Elle a donc décidé, au début de l'année 2025, de mettre en place **une mission d'information pour analyser les liens entre l'IA et la création artistique.**

Confiée aux sénatrices Laure Darcos et Agnès Evren et au sénateur Pierre Ouzoulias, cette mission d'information a organisé une quarantaine d'auditions et débattu avec une centaine d'interlocuteurs, parmi lesquels de nombreux représentants des ayants droit culturels, toutes filières confondues, des acteurs du secteur de la tech, ainsi que des experts juridiques et économiques. Compte tenu de la dimension fortement européenne du dossier, elle a aussi tenu à se déplacer à Bruxelles pour échanger avec les principales parties prenantes.

À l'issue de son travail, **la mission s'est forgé la conviction que l'opposition entre IA et création artistique était non seulement stérile, mais également mortifère pour les deux secteurs. La France et l'Europe ont tout à gagner, non pas à s'inscrire dans les pas d'autres puissances devenues peu amicales, mais à profiter de leurs atouts, au premier rang desquels la qualité et la diversité de leurs contenus culturels, pour ouvrir une réelle troisième voie de l'IA, respectueuse des droits et inspiratrice pour la création.**

I. PETIT PRÉCIS HISTORIQUE ET TECHNIQUE DE L'INTELLIGENCE ARTIFICIELLE

A. DES AUTOMATES ANTHROPOMORPHES DE L'ANTIQUITÉ À L'AGENT CONVERSATIONNEL CHATGPT : L'AVENTURE DE L'INTELLIGENCE ARTIFICIELLE

1. Aux origines imaginaires et rationnelles de l'intelligence artificielle

Si l'IA en tant que discipline scientifique et technologie est née au milieu du XX^e siècle, elle puise **ses racines à la fois dans la mythologie antique, l'histoire des sciences et la littérature**. L'idée de construire une intelligence, qui n'est pas d'origine humaine, fascine depuis des milliers d'années.

Les mythes antiques, grecs particulièrement, sont riches de créatures imaginaires douées d'attributs humains, voire de raison. Le personnage de Talos, dont l'un des mythes raconte qu'il a été créé à la demande de Zeus par Héphaïstos, dieu de la forge, était un immense automate de bronze ayant pour mission de protéger Europe - la mère du roi de Crète Minos - des envahisseurs, pirates et autres ennemis. Considéré comme l'une des premières manifestations de l'idée de robot de l'Histoire, Talos pouvait faire chauffer son corps de bronze dans le feu pour ensuite étreindre ses adversaires jusqu'à ce qu'ils périssent brûlés. Dans *l'Illiade* d'Homère, de nombreux objets ou créatures agissent par eux-mêmes : les navires des Phéaciens se pilotent de manière autonome, des trépieds se mettent en mouvement pour servir le vin aux dieux de l'Olympe, des servantes taillées en or par Héphaïstos sont dotées d'une conscience et anticipent les besoins de leur maître. Ainsi que le montrent les travaux d'Adrienne Mayor¹, chercheuse américaine en lettres classiques et en histoire des sciences, **les mythes antiques ont ainsi posé, de manière visionnaire, les interrogations éthiques qui émergent, des siècles plus tard, avec l'accélération du progrès technique et qui trouvent une nouvelle actualité avec l'essor de l'IA** : la puissance de la machine, la quête d'immortalité, le risque d'hubris...

Au-delà de ce vivier mythologique, **la future IA se nourrit du développement, au fil des siècles, des sciences, en tout premier lieu des mathématiques et, au sein de celles-ci, de l'algorithmique**. Le mot algorithme vient de la latinisation du nom d'un mathématicien perse du IX^e siècle, Al-Khwârizmî, considéré comme le père de l'algèbre. L'origine des algorithmes est toutefois bien plus ancienne puisque des procédés algorithmiques ont été retrouvés sur des tablettes écrites en cunéiforme par les Babyloniens au III^e millénaire avant J.C. et que des formules algorithmiques ont été

¹ Adrienne Mayor, "Gods and Robots", Princeton University Press, november 2018.

développées entre 300 et 200 avant notre ère par les mathématiciens grecs Euclide, Archimède et Ératosthène.

Qu'est-ce qu'un algorithme ? Il s'agit d'une suite finie et non ambiguë d'opérations ou d'instructions qui, à partir de données fournies en entrée (entrants ou *inputs*), permet d'obtenir un résultat sortant (ou *output*).

De manière imagée, un algorithme s'apparente à **une recette de cuisine** où les ingrédients, en suivant des étapes successives, permettent d'obtenir un plat.

Au XVII^e siècle, le **philosophe et mathématicien allemand Gottfried Leibniz apporte une contribution notable au développement de la pensée algorithmique**, qui préfigure les fondements de l'informatique moderne. Convaincu que mathématiques et pensée métaphysique ne font qu'un, il théorise le *calculus ratiocinator*, algorithme ou machine calculatoire permettant de démêler le vrai du faux dans toute discussion dont les termes seraient exprimés dans une langue philosophique universelle. Leibniz est aussi le concepteur d'un prototype de machine à calculer, capable d'effectuer les quatre opérations de l'arithmétique.

Au cours du XIX^e siècle, de nouvelles machines à calculer, ancêtres mécaniques des ordinateurs, sont mises au point par des scientifiques britanniques, notamment celle du mathématicien Charles Babbage sur laquelle sa compatriote Ada Lovelace développe le premier programme informatique de l'Histoire.

Parallèlement à ces progrès scientifiques, **la littérature du XIX^e siècle**, sous l'effet des bouleversements technologiques provoqués par la révolution industrielle, **se passionne pour les formes d'hybridation entre l'homme et la machine** et interroge la capacité de cette dernière à développer une conscience, **contribuant ainsi à dessiner l'avenir de l'IA**. Des ouvrages de science-fiction comme *Frankenstein* de Mary Shelley ou les romans de Jules Verne forgent l'imaginaire collectif et ouvrent la voie aux œuvres d'anticipation du siècle suivant, particulièrement dans le Septième Art.

2. Naissance et développement de l'intelligence artificielle : une histoire non linéaire

a) Les années 1950 : l'époque des pionniers de l'intelligence artificielle

Ce n'est que dans la première moitié du XX^e siècle que les avancées réalisées dans les domaines de la logique formelle et de l'informatique permettent de poser les premiers jalons de l'IA telle que la notion émergera dans les années 1950.

Considéré comme l'un des pères fondateurs de l'informatique, le mathématicien britannique Alan Turing publie en 1936 un article, *On Computable Numbers* (« *De la calculabilité des nombres* »), dans lequel il imagine

une machine dotée d'une bande de papier équipée d'une tête d'écriture et de lecture, pouvant théoriquement réaliser n'importe quel type de calcul, ce qui la rendait, d'après son concepteur, universelle. Cette « machine de Turing », comme elle sera nommée plus tard, préfigure le fonctionnement théorique des ordinateurs modernes.

Après la Seconde Guerre mondiale, Turing envisage progressivement la possibilité de construire une machine pouvant développer une forme d'intelligence grâce à une méthode d'entraînement. **En 1950, il fait publier dans la revue de philosophie *Mind* son article le plus célèbre, souvent considéré comme l'acte fondateur de l'idée d'IA, *Computing Machinery and Intelligence* (« *Les ordinateurs et l'intelligence* »).** Il y propose un test, qu'il nomme *imitation game* (*jeu de l'imitation*), désormais connu sous le nom de « **test de Turing** », dont le principe consiste à mettre en confrontation verbale un humain avec une machine imitant la conversation humaine et un autre humain. Dans le cas où l'homme qui engage la conversation n'est pas capable de dire lequel de ses interlocuteurs est une machine, on peut considérer que cette dernière a passé le test avec succès et qu'elle peut donc être qualifiée d'intelligente. Turing fait alors le pari que les machines vont réussir son test à moyen terme (« *d'ici à cinquante ans, il n'y aura plus moyen de distinguer les réponses données par un homme ou un ordinateur, et ce sur n'importe quel sujet* »).

Aujourd'hui encore, le test de Turing est régulièrement utilisé pour évaluer les IA contemporaines. En février 2024, trois chercheurs américains de l'université Stanford ont ainsi conclu que ChatGPT, l'agent conversationnel d'OpenAI, « *présente des traits de comportement et de personnalité qu'il est statistiquement impossible de distinguer d'un humain* ».

C'est à l'été 1956, lors d'une conférence organisée au Collège de Dartmouth dans le New Hampshire, que l'expression « intelligence artificielle » est officiellement utilisée pour définir un nouveau domaine de recherche. Les instigateurs de cette conférence, parmi lesquels figurent des personnalités scientifiques éminentes telles que John McCarthy et Marvin Minsky, sont animés par une conviction forte : celle de **la faisabilité de construire des machines capables de simuler les capacités cognitives humaines**. En créant les conditions d'un dialogue interdisciplinaire, où les connaissances issues de l'informatique, de la psychologie, de la philosophie, des neurosciences et des mathématiques pourraient s'enrichir mutuellement, ils aspirent à jeter les bases d'un domaine de recherche qui, jusqu'alors, n'a pas de contours clairement définis.

Véritable acte de naissance de l'IA, la Conférence de Dartmouth pose ainsi le présupposé selon lequel « *chaque aspect de l'apprentissage ou toute autre caractéristique de l'intelligence artificielle peut en principe être décrit avec une telle précision qu'une machine peut être fabriquée pour le simuler* ». **L'IA devient donc la science et l'ingénierie de fabrication de machines pouvant simuler tel ou tel aspect de l'intelligence humaine.** Il est intéressant de noter que, dans cette

acceptation, l'IA est à la fois une discipline scientifique et un savoir-faire pratique.

En 1959, le terme *machine learning* (apprentissage automatique) apparaît pour la première fois, utilisé par Arthur Samuel pour son programme capable d'apprendre à jouer aux dames au fil des parties. Cette technique, qui **permet aux algorithmes d'apprendre ou d'améliorer leurs performances en fonction des données qu'ils reçoivent**, devient une sous-branche à part entière de l'IA.

Définitions de l'intelligence artificielle

Il n'existe pas de définition unique de l'IA, mais le terme désigne généralement la **capacité de machines à effectuer des tâches associées à l'intelligence humaine**, comme l'apprentissage, le raisonnement, la résolution de problème, la perception ou la prise de décision. L'IA désigne également le **domaine de recherche** qui s'intéresse au développement de telles machines et la **technologie** qui permet de les fabriquer.

Définition de John MacCarthy : l'IA est « *la science et l'ingénierie de la fabrication de machines intelligentes* ».

Définition du Parlement européen : l'IA représente tout outil utilisé par une machine afin de « *reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité* ».

Définition du Larousse : l'IA est « *un ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine* ».

b) Les « printemps » et les « hivers » de l'intelligence artificielle

L'engouement suscité par la Conférence de Dartmouth ouvre **une période d'âge d'or pour l'IA, qui court jusqu'au début des années 1970**. Ce « premier printemps » se caractérise par un fort optimisme et des objectifs ambitieux, l'IA étant perçue comme un domaine de recherche très prometteur.

Portés par un soutien financier important tant de la part des pouvoirs publics (par exemple, l'agence américaine pour les projets de recherche avancée de défense) que des entreprises privées (par exemple, IBM), les chercheurs explorent l'IA dite « symbolique ». Celle-ci repose sur l'idée que la logique mathématique peut représenter, au moyen de symboles, des connaissances et modéliser des raisonnements. C'est au cours de cette première période qu'**une première concrétisation des réseaux de neurones artificiels voit le jour sous la forme du Perceptron** de Frank Rosenblatt et que **le premier agent conversationnel (chatbot)**, baptisé Eliza, est mis au point par Joseph Weizenbaum.

À l'enthousiasme des deux premières décennies succède **une phase de stagnation, qui s'étend de 1970 à 1980**, marquée par une prise de conscience croissante des défis inhérents à la réalisation des ambitions initiales de l'IA et une baisse des financements publics et privés. Les recherches se

concentrent alors sur la programmation logique et la représentation des connaissances. Cette décennie moins dynamique, connue sous le nom de « premier hiver de l'IA » en référence au contexte de la guerre froide, témoigne de **la cyclicité de son histoire, qui est faite de pics et de creux d'investissements et de confiance.**

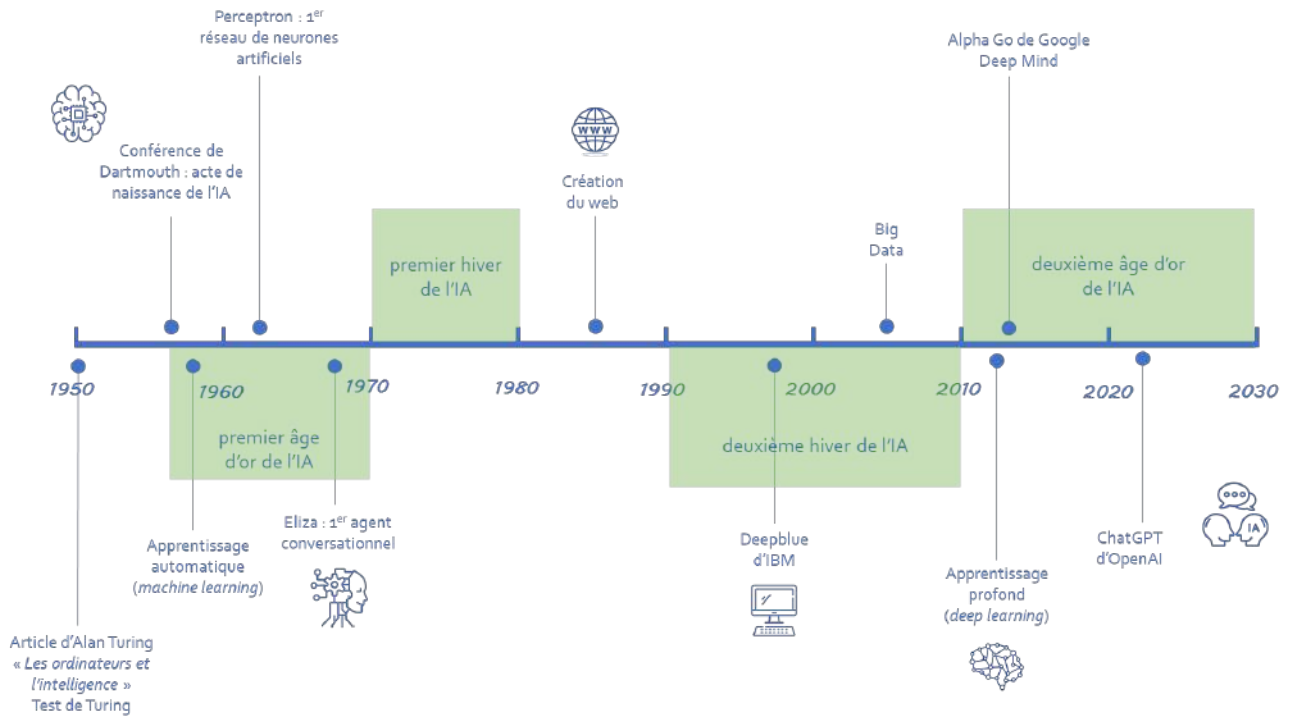
Au cours des **années 1980**, l'IA bénéficie d'un **regain d'intérêt** sous l'effet de nouveaux investissements publics de la part des États-Unis, de l'Europe et du Japon. Cette période voit **le développement des systèmes experts**, c'est-à-dire des programmes conçus pour imiter l'expertise humaine dans des domaines spécifiques et capables de résoudre des problèmes complexes. Même si cette nouvelle approche montre le potentiel pratique de l'IA, ses succès restent très relatifs car cantonnés à des domaines trop restreints et spécialisés.

Ce bilan en demi-teinte annonce **un « second hiver de l'IA » dans les années 1990**, celle-ci n'étant de nouveau plus une priorité des pouvoirs publics. Un événement majeur marque tout de même la décennie : en 1997, le système d'expert *Deep Blue* d'IBM bat le champion du monde d'échecs Garry Kasparov. Pour la première fois, une intelligence artificielle est capable de l'emporter sur une intelligence humaine.

Il faut attendre **les années 2010** pour que l'IA connaisse **un nouvel âge d'or ou « second printemps », cette fois-ci spectaculaire.** Le saut qualitatif qu'elle accomplit durant cette décennie est rendu possible par un contexte très porteur, qui se dessine à partir des années 2000. Avec l'apparition d'Internet puis le développement exponentiel de ses usages, le volume de données disponibles en ligne explose ; cette massification des données numériques (*big data*) constitue un formidable tremplin pour l'IA qui a besoin de « carburant » pour progresser. Parallèlement, les avancées technologiques en informatique permettent d'augmenter considérablement les capacités de calcul des ordinateurs. **La combinaison entre la masse de données accessibles et la puissance des machines permet de faire décoller l'apprentissage profond (*deep learning*),** grâce notamment aux travaux des scientifiques américains Yoshua Bengio, Geoffrey Hinton et du chercheur français Yann LeCun. Cette **technique d'apprentissage basée sur des réseaux de neurones artificiels** (voir *infra*), sous-discipline du *machine learning*, permet de réaliser des avancées très significatives en matière de reconnaissance vocale, de traitement du langage naturel (*NLP*), de reconnaissance visuelle et d'apprentissage par renforcement. En 2015, le programme AlphaGo, mis au point par la société Google DeepMind et qui a appris à jouer au jeu de go par le biais du *deep learning*, bat le champion européen Fan Hui par cinq parties à zéro.

En 2017, l'IA franchit **une nouvelle étape charnière** avec la montée en puissance de l'IA dite « **générative** », elle-même issue du *deep learning* et rendue possible par l'invention cette année-là de **la technologie Transformer** (cf. *infra*). C'est de cette technologie que naissent les grands modèles de

langage ou LLM (*Large Language Model*), dont l'exemple le plus célèbre est **ChatGPT**, lancé en **2022** par l'entreprise américaine OpenAI.

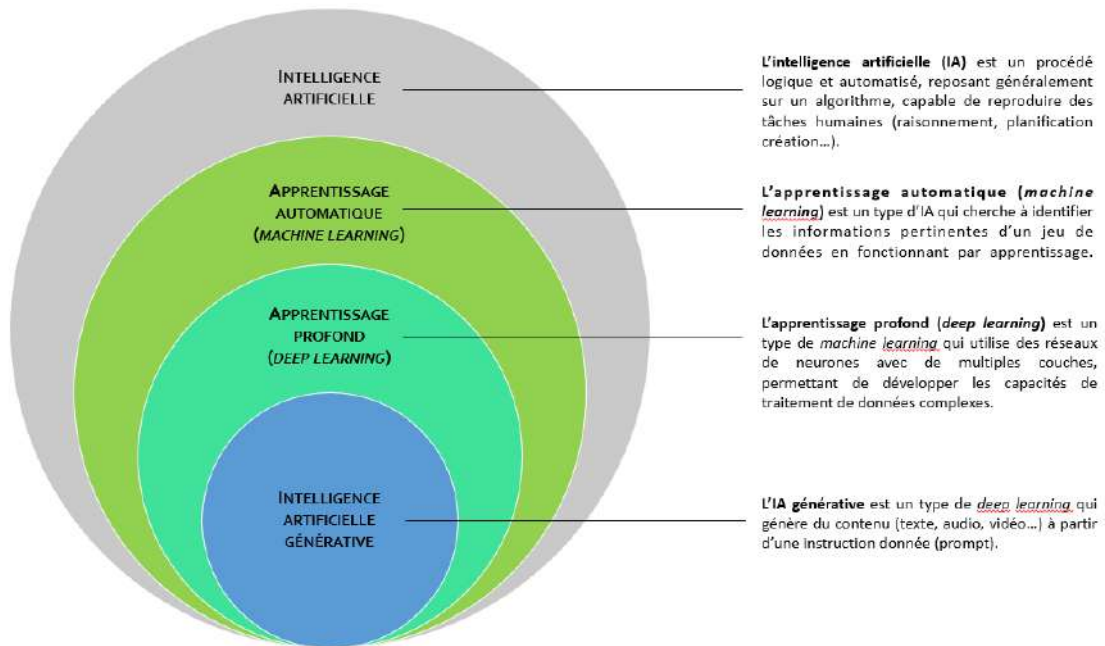


B. GLOSSAIRE DE L'INTELLIGENCE ARTIFICIELLE

Le vocabulaire autour de l'IA est composé d'une multitude de termes et d'expressions, souvent techniques et pas toujours utilisés à bon escient. L'objectif ici est de faciliter la compréhension des différentes notions relatives à l'IA, en identifiant ce qui les lie et les différencie.

1. L'intelligence artificielle et ses sous-domaines

Au fur et à mesure de ses avancées scientifiques et technologiques, l'IA, qui est en quelque sorte la **discipline-mère**, a donné naissance à des **sous-domaines**, qui peuvent être représentés en sous-ensembles concentriques.



a) L'apprentissage automatique (*machine learning*)

Le *machine learning* est un champ d'études au sein de l'IA qui vise à **donner aux machines la capacité d'apprendre à partir de données, via des modèles mathématiques utilisant des algorithmes probabilistes.**

Plus précisément, il s'agit du procédé par lequel une machine, au moyen de techniques algorithmiques et sans l'intervention additionnelle de son programmeur, traite et analyse des données, apprend de ces données, estime la probabilité de résultats et prend des décisions fondées sur ce qu'elle a appris. **Plus la machine réalise cet apprentissage, plus ses performances s'améliorent.** En outre, **plus le volume de données auquel elle a accès est important, plus ses résultats progressent.**

Schéma d'un modèle d'apprentissage automatique (*machine learning*)



Le *machine learning* se décline en **quatre principaux modèles d'apprentissage**, qui emploient chacun des techniques algorithmiques différentes :

- **l'apprentissage supervisé** : dans ce premier modèle, la machine s'entraîne, au moyen d'un algorithme, à une tâche déterminée (ou clef de réponse) en utilisant un jeu de données assorties chacune d'une annotation (ou données étiquetées) indiquant le résultat attendu ;

- **l'apprentissage non supervisé** : dans ce deuxième modèle, la machine s'entraîne, au moyen d'un algorithme, sans tâche déterminée (sans clef de réponse) en utilisant un jeu de données brutes (sans étiquetage). En détectant des récurrences, des similarités, des corrélations entre certaines de ces données, elle obtient un résultat ;

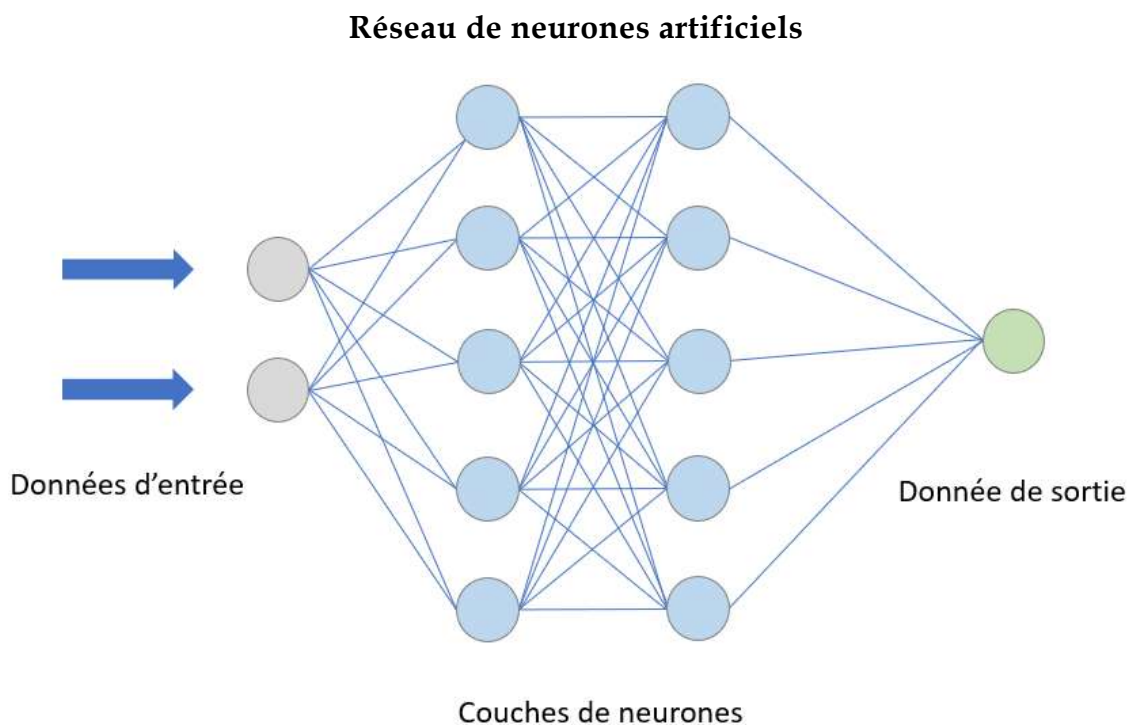
- **l'apprentissage semi-supervisé** : dans ce troisième modèle, intermédiaire entre les deux premiers, la machine s'entraîne, au moyen d'un algorithme, à partir de petits volumes de données étiquetées pour analyser de gros volumes de données non étiquetées ;

• **L'apprentissage par renforcement** : dans ce quatrième modèle, la machine, qui connaît la clef de réponse, apprend en expérimentant différentes actions et en attribuant une valeur positive ou négative à chacune d'entre elles en fonction du résultat obtenu. Elle cherche, au fil des expériences, à maximiser la somme des récompenses, donc à parvenir à une solution décisionnelle optimale.

Dans chacun de ces modèles, **une ou plusieurs techniques algorithmiques** peuvent être appliquées ; tout dépend des jeux de données qui seront utilisés et de l'objectif visé au niveau des résultats. Par nature, les algorithmes du *machine learning* sont conçus pour classer des éléments, repérer des motifs récurrents (*patterns*), prévoir des résultats (analyse probabiliste) et prendre des décisions éclairées. Les algorithmes peuvent être mis en œuvre individuellement ou en groupe dans le but d'atteindre la plus grande précision possible lorsque les données utilisées sont complexes et imprévisibles.

b) L'apprentissage profond (deep learning)

Le *deep learning* est un **sous-ensemble du machine learning** qui recourt à l'apprentissage supervisé, mais avec **une architecture bien spécifique, celle d'un réseau de neurones artificiels** agencé sous la même forme que les neurones d'un cerveau biologique. Les neurones artificiels composant un réseau s'appellent des « nœuds », qui sont connectés et regroupés en « couches ». Lorsqu'un nœud reçoit un signal digital, il le transmet alors à d'autres neurones appropriés, qui fonctionnent en parallèle.



Cet apprentissage, qui **fait appel à d'énormes volumes de données, complexes et disparates**, est **qualifié de « profond » (*deep*) car il fonctionne par « couches »** : les résultats de la première « couche » de neurones vont servir d'entrée au calcul des autres, et ainsi de suite. Plus on augmente le nombre de couches, plus les réseaux de neurones apprennent des choses compliquées, abstraites, se rapprochant de plus en plus du fonctionnement d'un cerveau humain.

Par exemple, un modèle d'apprentissage profond qui traite des images de la nature et recherche des roses jaunes va d'abord reconnaître une plante dans la première couche. Au fur et à mesure qu'il va avancer dans les couches, il va ensuite identifier une fleur, puis une rose, et enfin une rose jaune.

Il existe **différents types de réseaux neuronaux artificiels**, parmi lesquels :

- **les réseaux neuronaux convolutifs (CNN)** : ils sont basés sur la structure du cortex visuel et utilisent des couches convolutives pour extraire les caractéristiques d'une image d'entrée. Ces CNN sont principalement utilisés pour des tâches de reconnaissance d'images ;

- **les réseaux neuronaux récurrents (RNN)** : ils permettent de « se souvenir » des informations passées et de les utiliser pour prendre des décisions en temps réel. Ces RNN sont spécialement utilisés pour l'analyse de séquences de données ;

- **les réseaux antagonistes génératifs (GAN)** : introduite pour la première fois en 2014, cette architecture repose sur deux réseaux neuronaux qui s'affrontent. L'un des réseaux, le générateur, crée des données originales, tandis que l'autre, le discriminateur, évalue si les données sont générées par l'IA ou réelles. Grâce à une méthode d'apprentissage profond et à une boucle de rétroaction qui pénalise le discriminateur en cas d'erreur, les GAN apprennent à générer un contenu de plus en plus réaliste ;

- **les réseaux neuronaux Transformer (TNN)** : cette architecture de réseau neuronal, introduite par l'article « Attention is All You Need » publié par des chercheurs de Google en 2017, est très novatrice par rapport aux architectures traditionnelles de traitement de séquences de données comme les CNN ou les RNN. Transformer repose sur un mécanisme d'auto-attention : au lieu de traiter les données dans l'ordre, il examine simultanément différentes parties de la séquence et détermine lesquelles sont les plus importantes. Ce fonctionnement non séquentiel, qui apporte flexibilité, adaptation et rapidité, permet de réaliser une multitude de tâches (génération d'images, traduction automatique, compréhension des séquences, détection d'anomalies...).

c) *L'intelligence artificielle générative et sa capacité créatrice*

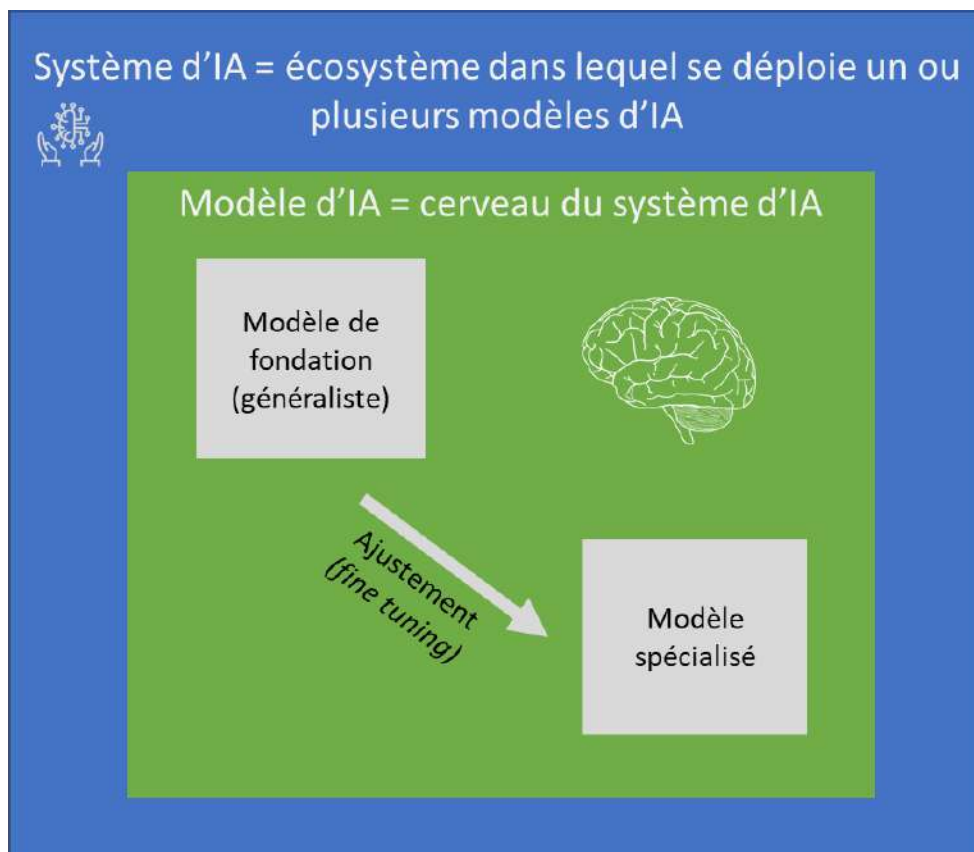
L'IA dite « générative » (qualificatif dérivé du supin latin *generatum*, qui signifie « pour créer ») est un type d'IA capable de **générer des contenus nouveaux couvrant un spectre très large** (texte, code informatique, images, musique, audio, vidéos, etc.), à partir des données grâce auxquelles elle a été formée. Ces contenus nouveaux ressemblent à ce que l'on peut trouver dans ces données dites « d'entraînement », mais ils ne sont pas semblables, d'où leur caractère original.

C'est cette créativité – entendue au sens de la **capacité à générer une production originale** - qui distingue l'IA générative de l'IA prédictive, laquelle prévoit et anticipe mais ne crée pas.

L'IA générative repose principalement sur des techniques d'apprentissage profond. Autrement dit, elle est **une application du *deep learning*** qui se concentre sur la génération de contenus nouveaux.

L'IA générative, tout comme l'apprentissage automatique dont elle est issue, est par nature **probabiliste** : à partir des données d'entraînement, elle va estimer la probabilité de différents résultats et générer des contenus sortants qui découlent des probabilités apprises. Cela explique pourquoi **elle est susceptible de produire des résultats erronés ou de pures inventions**, appelées hallucinations, qui sont à l'origine de sa dégénérescence (voir *infra*).

2. Modèle d'intelligence artificielle vs système d'intelligence artificielle



Un **modèle d'IA** est un programme informatique qui applique un ou plusieurs algorithmes à un jeu de données pour reconnaître des motifs, réaliser des tâches spécifiques ou prendre des décisions. **Un modèle d'IA est intégré à un système d'IA, dont il est une composante essentielle, son cerveau** en quelque sorte. Il est construit et optimisé à travers un processus d'entraînement, où il apprend à partir de données entrantes pour améliorer sa précision et son efficacité.

Un **système d'IA** est une application **plus large** qui englobe non seulement un ou plusieurs modèles d'IA, mais aussi **tout l'écosystème** nécessaire à leur fonctionnement. Il inclut des processus de collecte de données, des interfaces utilisateurs, une infrastructure technique (serveurs, bases de données) pour déployer efficacement l'IA dans des environnements pratiques. En d'autres termes, un système IA est une **solution complète qui met en œuvre des modèles IA dans un cadre opérationnel**.

Au sein des modèles d'IA, on distingue **les modèles de fondation**, à usage général, et **les modèles spécialisés** :

- un modèle de fondation désigne un modèle d'IA de très grande taille, entraîné sur d'énormes quantités de données brutes (non étiquetées), dont les capacités sont générales, et qui peut être adapté à une grande diversité de tâches distinctes (compréhension du langage, génération de textes et d'images, conversation en langage naturel...), notamment après un ajustement (*fine-tuning*) supplémentaire. Cette flexibilité, combinée à la capacité à apprendre à partir de vastes ensembles de données, le rend particulièrement puissant. Un modèle de fondation peut être adapté à des tâches non génératives, comme la classification, l'analyse, la prédiction, ou à des tâches génératives ; il est donc plus large dans sa conception et son application potentielle qu'un modèle d'IA générative ;

- à l'inverse, **un modèle qui est d'emblée spécialisé**, sans passer par l'ajustement d'un modèle de fondation, n'est dédié qu'à **une seule fonction** ou appliqué qu'à **un seul domaine**. Un modèle spécialisé n'est toutefois pas forcément de petite taille.

C. LA CONCEPTION ET LE FONCTIONNEMENT D'UN MODÈLE D'INTELLIGENCE ARTIFICIELLE GÉNÉRATIVE

1. Une diversité de modèles et d'acteurs

Un **modèle d'IA générative** est un modèle d'IA qui a **la capacité de générer un contenu nouveau et original** (un texte, une image, une vidéo...), en se basant sur ce qu'il a appris. Un modèle d'IA générative n'est **pas forcément un modèle de fondation** (voir *supra*).

Il existe plusieurs types de modèles d'IA générative :

- un **modèle de langage** est un modèle d'IA générative **capable de générer un contenu textuel**. On parle de **grand modèle de langage** (*Large Language Model* ou **LLM**) lorsque le modèle possède un grand nombre de paramètres. Les LLM sont les modèles d'IA générative les plus courants aujourd'hui ; ChatGPT est le plus connu d'entre eux ;

- un **modèle de diffusion** est un modèle d'IA générative capable de générer un **contenu visuel**. C'est le cas, par exemple, de DALL-E et de Midjourney ;

- un **modèle multi-modal** est un modèle d'IA générative capable de générer **un contenu sous différents formats** (texte, image, son, vidéo...) ; GPT-4 en est.

Différents acteurs interviennent dans la chaîne de développement et d'utilisation d'un modèle d'IA, parmi lesquels :

- le « **fournisseur** » : selon la définition qu'en donne le règlement européen établissant des règles harmonisées concernant l'intelligence artificielle (RIA) du 13 juin 2024, il s'agit d' « *une personne physique ou morale, une autorité publique, une agence ou un autre organisme qui développe ou fait développer un système d'IA ou un modèle d'IA à usage général et le met sur le marché ou met le système d'IA en service sous son propre nom ou sa propre marque, que ce soit à titre onéreux ou gratuit* ».

→ **Exemples de fournisseurs d'IA** : OpenAI, Google, Amazon, Microsoft, IBM, Mistral AI...

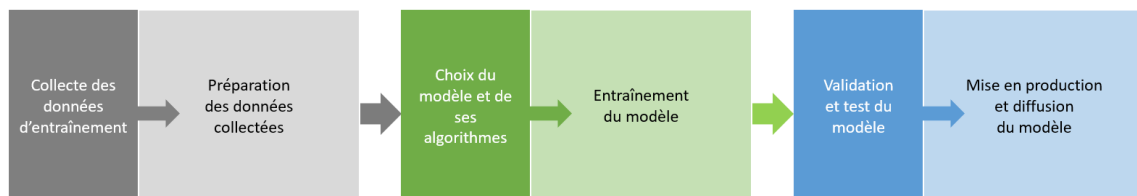
- le « **développeur** » : il s'agit du professionnel capable de concevoir un modèle d'IA. Ses missions incluent l'analyse des besoins, la collecte et la préparation des données, le choix de la technique d'apprentissage, l'entraînement du modèle, son évaluation, son intégration dans une application ou un système, sa surveillance...

- le « **déploieur** » : selon la définition du RIA, il s'agit de « toute personne physique ou morale, autorité publique, agence ou autre organisme qui utilise un système d'IA sous sa responsabilité, sauf si l'utilisation est à des fins personnelles non professionnelles ».

→ **exemples de déploieurs d'IA** : une banque qui a recours à une IA pour analyser les dossiers de prêt, un cabinet RH qui déploie une IA pour trier automatiquement des CV, une entreprise de cybersécurité qui utilise l'IA pour détecter des fraudes, un organisme public qui met en place un *chatbot* (assistant virtuel de conversation) automatisé pour répondre aux citoyens...

2. Les différentes étapes de la conception d'un modèle d'intelligence artificielle générative

La chaîne de conception d'un modèle d'IA générative comprend plusieurs étapes :



a) La collecte des données d'entraînement ou le grand moissonnage

Les données d'entrée (aussi dénommées données brutes) sont **un élément-clef** pour la création d'un modèle d'IA : ce sont en quelque sorte **les ingrédients** sans lesquels « la recette » IA ne peut être cuisinée. **Sans données, une IA reste inerte, incapable d'apprendre et encore moins de prendre des décisions pertinentes.**

Le terme « données » revêt ici un sens générique : il peut s'agir de données textuelles, de données d'image, de données sonores, de données vidéo, ou de données multi-supports...

Pour garantir l'efficacité, la précision et la fiabilité du modèle d'IA, **les données doivent répondre à plusieurs critères :**

- elles doivent être **en quantité suffisante** : pour qu'un modèle soit bien entraîné, il a besoin d'analyser beaucoup d'exemples, donc un volume très important de données ;
- elles doivent être **de qualité** : sans données pertinentes, précises et fiables, même les modèles d'IA les plus sophistiqués peuvent générer des résultats biaisés ou erronés ;
- elles doivent être **diverses** : plus les données sont variées, plus le modèle améliore la précision et la pertinence de ses réponses.

Il existe **différents canaux** de collecte de données d'entraînement.

La **plus grande source d'approvisionnement** est constituée de **données accessibles publiquement en ligne**. Le fait que ces données soient « publiques » ne signifie pas qu'elles soient librement exploitables - certaines peuvent être protégées par des droits, comme le droit de la propriété intellectuelle s'il s'agit d'un contenu considéré comme « original » (texte, image, vidéo...) ou le droit au respect de la vie privée s'il s'agit de données à caractère personnel (nom, image ou vidéo identifiant une personne physique), ni qu'elles soient licites - certaines peuvent avoir été piratées.

L'accès à ces données publiques se fait *via* des moteurs de recherche généralistes comme Google ou Yahoo, des bases de données spécialement constituées pour servir à alimenter les modèles d'IA comme Laion, des bases de données spécialisées dans certains domaines comme Library Genesis (LibGen) pour les contenus scientifiques, ou la base de données de l'organisation Common Crawl qui constitue une sorte de répertoire recensant tous les contenus trouvés sur Internet depuis une dizaine d'années¹.

Pour collecter massivement et automatiquement des données librement accessibles en ligne, il est fait recours à **la technique du *web scraping* qui signifie littéralement « raclage », mais qui est plus couramment traduite par le terme « moissonnage »**. Elle permet, à l'aide de *bots* (applications logicielles automatisées) fonctionnant comme des robots « racleurs » ou « moissonneurs », de récupérer des données à partir du code source des pages Web concernées, puis de les restituer dans une base de données organisées de type tableur. De manière imaginée, cette opération revient à envoyer un robot dans une librairie pour qu'il rapporte des livres spécifiques. Dans le cas d'espèce, la librairie est un site Web et les livres sont des données.

Bien que **les pratiques de moissonnage** offrent des avantages importants en termes de rapidité, de quantité et de diversité des données collectées, elles posent **de nombreuses questions sur la protection de la vie privée, le respect de la propriété intellectuelle, la rétribution des ayants droit**. La Cnil a d'ailleurs régulièrement appelé à la vigilance concernant ces pratiques, dont la généralisation opère selon elle *« un changement de nature quant à l'utilisation d'Internet, dans la mesure où toutes les données publiées en ligne par une personne sont désormais susceptibles d'être lues, collectées et réutilisées par des tiers, ce qui constitue un risque important et inédit pour les personnes »*. En conséquence, la Cnil recommande la mise en place d'un encadrement législatif *ad hoc*², qui permettrait à la fois de sécuriser les organismes y ayant recours, de réguler cette pratique et de protéger les données librement accessibles en ligne.

Il existe **trois autres sources d'approvisionnement en données d'entraînement**, qui sont d'un accès moins facile que la première :

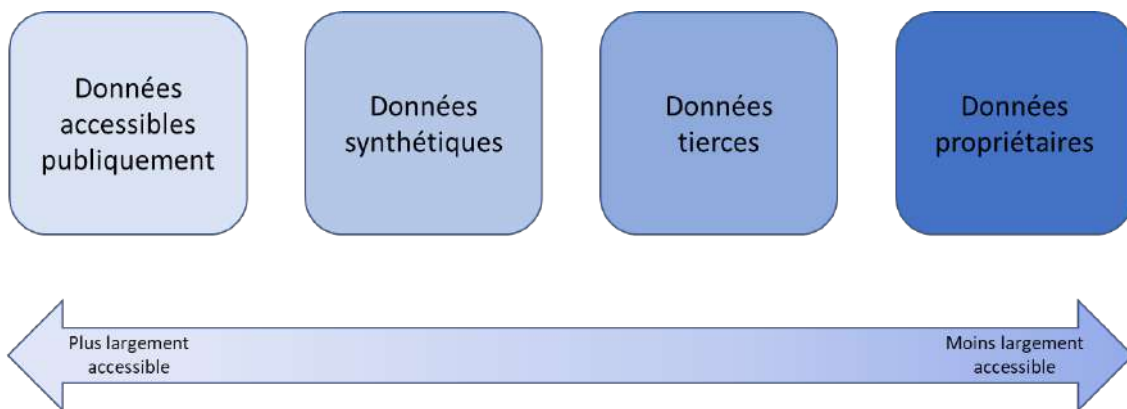
- **les données synthétiques**, c'est-à-dire des données artificiellement conçues pour imiter les données du monde réel. Elles sont générées par des méthodes statistiques ou par des techniques d'IA ;

¹ Common Crawl est une organisation à but non lucratif américaine, créée en 2017, qui explore le Web et fournit gratuitement ses archives et ses ensembles de données au public. Par des opérations d'extraction mensuelles qui capturent des milliards de pages Web, la base de données de Common Crawl est devenue une source indispensable pour l'entraînement des modèles d'IA.

² Avis de la Cnil du 15 décembre 2022 sur le projet « Polygraphe ».

- **les données propriétaires de tiers**, à savoir des données collectées par des entités externes comme des courtiers en données. Les principales entreprises de ce secteur sont Experian, Equifax, Acxiom et Epsilon ;
- **les données internes** qui sont directement détenues par les entreprises développant des modèles d'IA.

Dans son récent rapport sur la rémunération des contenus culturels utilisés par les systèmes d'IA¹ commandé par le Conseil supérieur de la propriété littéraire et artistique (CSPLA), la professeure Joëlle Farchy résume le degré d'accessibilité des différentes sources de données d'entraînement par le schéma suivant :



b) La préparation des données collectées

Avant d'être utilisées, les données brutes collectées doivent être préparées. Elles peuvent en effet comporter des anomalies, des incohérences, des doublons ou des valeurs manquantes. Une préparation rigoureuse permet de corriger ces problèmes qui pourraient entraîner des imprécisions, des biais, des erreurs, voire des hallucinations.

Cette étape comprend plusieurs tâches successives :

- **le nettoyage des données** : il vise à éliminer les anomalies, à corriger les erreurs, à supprimer les doublons, à remplacer les valeurs manquantes ;
- **la structuration et la transformation des données** : une fois nettoyées, les données doivent être préparées pour s'adapter aux exigences des algorithmes d'apprentissage. C'est là qu'intervient le processus dit de « **tokenisation** », par lequel **les données sont segmentées en unités plus petites appelées tokens**, plus faciles à appréhender par le modèle. S'en suit un processus d'encodage ou de vectorisation, au cours duquel chaque *token* est converti en valeur numérique que le modèle pourra directement traiter ;

¹ « Rémunération des contenus culturels utilisés par les systèmes d'intelligence artificielle », projet de rapport - volet économique, mission confiée par le Conseil supérieur de la propriété littéraire et artistique, Joëlle Farchy et Bastien Blain, mai 2025.

- **la normalisation et la mise à l'échelle des données** : les données peuvent présenter des écarts importants en termes de grandeur ou d'échelle, ce qui peut perturber certains algorithmes d'apprentissage. La normalisation et la mise à l'échelle permettent de les harmoniser en ajustant leur valeur à une plage standard ;

- **l'étiquetage des données** : cette étape consiste à associer une annotation spécifique à chaque donnée. Cet étiquetage sert de guide pour l'apprentissage des modèles et garantit que les données sont interprétées correctement pendant l'entraînement ;

- **l'enrichissement des données** : pour améliorer la pertinence et l'affinage des données brutes, des informations supplémentaires peuvent être ajoutées. Cet enrichissement inclut l'intégration de métadonnées (c'est-à-dire de données fournissant des informations sur d'autres données), l'ajout de contextes ou la combinaison avec des données externes complémentaires ;

- **l'équilibrage des données** : un jeu de données déséquilibré, où certaines catégories sont sur-représentées, peut introduire des biais dans les modèles d'IA. L'équilibrage consiste à ajuster la distribution des données en réduisant ou en augmentant certaines catégories de données par rapport à d'autres ;

- **la validation des données** : cette étape inclut des contrôles automatiques ou manuels pour détecter d'éventuelles anomalies restantes et vérifier que le jeu de données est conforme aux exigences du projet ;

- **le partitionnement des données** : la dernière étape de la préparation des données consiste à les diviser en trois ensembles distincts :

- ✓ **un premier ensemble**, comprenant généralement entre 70 % et 80 % des données, est consacré à l'entraînement du modèle ;
- ✓ **un deuxième ensemble**, composé entre 10 % à 15 % des données, à sa validation ;
- ✓ **un troisième ensemble**, également composé de 10 % à 15 % des données, à son test.

c) *L'entraînement du modèle à partir des données préparées*

Si la collecte et la préparation des données s'apparentent en quelque sorte à la base de la construction « IA », le choix de la technique d'entraînement (*machine learning*, *deep learning*) et des algorithmes associés constituent son architecture générale.

Une fois cette architecture mise en place, l'étape de l'entraînement peut commencer. Elle consiste à alimenter le modèle en données préparées, à examiner les résultats sortants (*outputs*), à procéder à des ajustements pour améliorer le modèle, et à reproduire ces mêmes étapes autant de fois que nécessaire.

L'entraînement d'un modèle d'IA est donc **un processus itératif, semblable à la manière dont un enfant acquiert une compétence par la répétition.**

S'il s'agit d'un modèle de fondation, **une étape supplémentaire** peut être introduite, consistant à procéder à **un ajustement (*fine-tuning*) pour le spécialiser.** Le modèle est alors réentraîné sur **des données d'affinage, spécialisées.**

d) La validation et le test du modèle

Une fois que le modèle a réussi la phase d'entraînement initiale, la phase de validation consiste à utiliser à **un jeu de données distinct** du jeu de données d'entraînement. L'objectif est d'évaluer les performances du modèle à partir de nouvelles données, généralement plus complexes.

Après la validation du modèle, un nouveau jeu de données est utilisé pour tester l'exactitude du modèle. Si le modèle fournit des résultats précis avec ces données de test, il est prêt à être mis en service. Si le modèle présente des lacunes, le processus d'entraînement se répète jusqu'à ce que le modèle respecte les objectifs de performance définis.

e) La mise en production et le déploiement du modèle

La mise en production, aussi appelée phase d'inférence, correspond au processus par lequel un modèle, préalablement entraîné, est utilisé pour effectuer des prévisions ou produire un résultat à partir de **nouvelles données** - aussi **appelées données « fraîches »** ou **« données d'ancrage »** - sans nécessiter d'entraînement. L'apport de ces nouvelles données **permet d'ancrer le modèle dans l'actualité ou dans un contexte bien spécifique.**

Une fois le modèle ancré, il est prêt à être intégré dans une infrastructure d'hébergement déjà existante ou nouvellement créée, de type solution *cloud*, installation sur site Web, application spécifique... Le modèle peut être libre d'accès (*open source*) ou distribué sous licence.

**Les trois grandes catégories de données utilisées
pour le fonctionnement des modèles d'IA**

Les données d'entraînement, « raclées » en très grande quantité (de l'ordre de plusieurs millions voire plusieurs milliards) pour servir à entraîner le modèle.

Les données d'affinage ou d'ajustement, sélectionnées pour spécialiser le modèle sur certaines tâches.

Les données d'ancrage, « fraîchement » recueillies pour ancrer le modèle dans l'actualité.

II. LA CRÉATION ARTISTIQUE FACE À LA VAGUE DE L'INTELLIGENCE ARTIFICIELLE : NOUVEL HORIZON ARTISTIQUE OU PARASITISME CRÉATIF ?

A. COMMENT L'INTELLIGENCE ARTIFICIELLE BOULEVERSE LE PROCESSUS DE CRÉATION

1. Des premières expériences artistiques recourant à l'intelligence artificielle aux premières œuvres intégralement générées par celle-ci

La création artistique, la science et la technique sont depuis toujours très imbriquées. **Les avancées scientifiques permettent de développer de nouvelles techniques que le secteur artistique, qui peut d'abord s'y montrer réticent voire hostile, s'approprie ensuite pour faire évoluer la création.**

L'invention de la photographie au début du XIX^e siècle est un exemple particulièrement caractéristique de cette ambivalence. Perçue initialement comme une menace par certains artistes, parmi lesquels les poètes Alphonse de Lamartine et Charles Baudelaire, qui craignaient qu'elle ne remplace l'art pictural, la photographie trouve progressivement sa place dans le monde de la création, y compris chez les peintres qui l'utilisent comme moyen de reproduction et de documentation de leur œuvre, à l'instar de Gustave Courbet ou d'Edgar Degas. Dans la deuxième moitié du XIX^e siècle, elle acquiert progressivement ses lettres de noblesse en tant qu'art avec les premiers photographes-artistes comme Gustave Le Gray ou Eugène Atget.

L'apparition des premiers ordinateurs un peu avant le milieu du XX^e siècle bouleverse à son tour la pratique artistique, en permettant **l'utilisation de techniques d'automatisation dans le processus créatif**. C'est à cette époque qu'émerge **l'art génératif**, mouvement artistique qui explore les possibilités offertes par l'informatique et les algorithmes pour **permettre à des machines de créer de manière autonome**. Le cybernéticien Albert Ducrocq est l'un des précurseurs dans ce domaine : il crée en 1953 l'un des premiers générateurs automatiques de poèmes, dénommé Calliope. Des expérimentations **dans les arts visuels** voient également le jour, comme celles de l'artiste **Jean Tinguely** qui développe une série de vingt Méta-Matic entre 1955 et 1960, sorte de sculptures animées capables de dessiner et de créer des œuvres d'art. Le **secteur musical** n'échappe pas à cette fascination pour l'électronique : en 1955, le premier morceau de musique généré par ordinateur est créé **par Lejaren Hiller et Leonard Isaacson** à l'Université de l'Illinois¹ ; en 1958, le compositeur français Pierre Schaeffer crée le groupe de recherches musicales (GRM), important centre de recherche et d'expérimentations en musique électroacoustique.

¹ La pièce se composait de quatre mouvements. Le premier créait la mélodie, le deuxième générait des segments à quatre voix en utilisation des règles stochastiques, le troisième abordait le rythme et la dynamique, le quatrième expérimentait des grammaires génératives probabilistes.

Toutes ces expériences artistiques préfigurent **l'arrivée de l'IA dans le monde de la création à partir des années 1960**. Les potentialités qu'elle offre sont notamment mises à profit par **les pionniers de l'art numérique**, comme Michael Noll, qui crée certaines des premières images générées par ordinateur, ou Vera Molnar, première artiste à produire en France des dessins numériques. En 1970, l'artiste Harold Cohen met au point un programme informatique, nommé Aaron, capable de créer des dessins et des peintures de manière autonome.

C'est dans la décennie 2010, **avec l'essor des techniques d'apprentissage profond et l'apparition de l'IA générative, qu'un changement de paradigme s'opère**. Jusqu'alors relativement limitées, **les capacités démiurgiques de l'IA connaissent un changement d'échelle sans précédent**. Les réseaux antagonistes génératifs (GAN), introduits en 2014, sont les premiers modèles d'IA générative dont se servent les artistes. Le lancement en 2022 de modèles tels Stable Diffusion, DALL-E et Midjourney marque un véritable tournant dans l'histoire de la création. Désormais, l'IA est en mesure de générer de nouvelles œuvres (dessin, morceau de musique, texte...) à partir des milliards de données qui l'ont nourrie, la part humaine de ces créations se limitant à la commande adressée à la machine (*prompt*). **Cette nouvelle donne technologique, qui peut être qualifiée de disruptive, soulève des interrogations fondamentales, d'ordre éthique et juridique**, sur l'essence même de l'acte de création, le degré de créativité d'une œuvre, la fonction de l'artiste et le devenir de l'expression artistique.

2. Une échelle de gradation de la place de l'intelligence artificielle dans le processus créatif

a) L'intelligence artificielle comme outil au service du créateur

Aujourd'hui présente dans toutes les esthétiques, l'IA offre aux artistes un arsenal d'outils novateurs, leur permettant de faciliter, d'enrichir, d'étendre leur pratique créative. **Cette assistance se manifeste de multiples façons**, à différentes étapes du processus créatif :

- comme outil d'inspiration, l'IA propose des idées, des suggestions d'amélioration, des associations ou des comparaisons avec d'autres approches artistiques, stimulant ainsi la créativité de l'artiste ;

- comme outil d'automatisation, elle prend en charge certains aspects techniques, permettant à l'artiste de se concentrer sur les étapes les plus créatives de son œuvre ;

- comme outil de perfectionnement, elle affine certaines caractéristiques matérielles d'une œuvre ;

- comme outil de diffusion, elle améliore l'accessibilité d'une œuvre.

Ces formes d'aide à la création ne se substituent pas à la vision créative de l'artiste, mais ouvrent plutôt de nouvelles perspectives d'expression et d'expérimentation.

b) L'intelligence artificielle comme co-partenaire de création

Au-delà de la simple assistance, l'usage de l'IA générative dans la création ouvre aussi la voie à **des pratiques collaboratives entre l'artiste et la machine.**

Cette co-crédation est rendue possible par la complémentarité des compétences : tandis que l'artiste apporte sa sensibilité esthétique, son intention artistique, sa compréhension du contexte culturel, historique et social dans lequel va s'inscrire l'œuvre, l'IA déploie sa capacité d'analyse et de traitement de grandes quantités de données. Des échanges entre ces deux pôles émerge **une collaboration hybride qui ouvre un vaste champ de possibilités créatives** et qui s'inscrit, selon certains spécialistes, dans la continuité du travail en atelier où la créativité s'exerce indirectement *via* des instructions, la réalisation étant confiée à des exécutants.

c) L'intelligence artificielle comme créateur à part entière ?

Avec l'essor et la démocratisation des techniques d'IA générative, créer une image, un morceau de musique, un texte, devient à la portée de tous, moyennant la maîtrise des *prompts*. La création artistique humaine, qui reposait jusqu'alors sur un subtil dosage entre des qualités innées et des compétences acquises, est aujourd'hui confrontée à **une nouvelle forme de création, accessible en quelques mots-clés.**

Cette situation pose avec une acuité particulière **la question du rôle de l'humain dans la création et de la potentielle substitution des artistes par l'IA.** Suscitant de vifs débats, celle-ci ne peut recevoir de réponse tranchée car nul ne peut présager des futures avancées technologiques de l'IA.

À ce jour, il peut simplement être constaté que **si l'IA est en capacité de produire une œuvre nouvelle, elle le fait sans intention, sans émotion, sans vécu, sans prise en compte du contexte.** Elle ne fait que répondre à une commande donnée par l'homme, à partir de contenus déjà existants, sans conscience du sens et de la portée de sa création. L'humain reste celui qui initie la démarche, qui choisit de donner tel sens et telle forme à l'œuvre. L'IA demeure donc une potentialité que l'homme, avec plus ou moins de talent, peut exploiter à des degrés divers.

3. Principaux cas d'usage de l'intelligence artificielle dans les industries culturelles et créatives

IA et création artistique : un sujet d'exposition

Preuve que le sujet fascine, le musée du Jeu de Paume à Paris propose actuellement une grande exposition « *Le monde selon l'IA* », la première de cette ampleur à explorer les interactions entre l'IA et la création artistique, en embrassant toute une variété de médiums, de la photographie au cinéma, en passant par la peinture, la sculpture et la vidéo.

Y est présentée une sélection d'œuvres créées entre 2016 et aujourd'hui, dont plusieurs inédites, qui posent la question de l'expérience du monde « selon l'IA » ou « au prisme de l'IA ».

a) Dans les arts visuels

De par leur rapport intrinsèque à l'image, **les arts visuels constituent un terrain artistique très propice à l'usage de l'IA**. C'est d'ailleurs dans les arts visuels que celle-ci a trouvé ses premières applications artistiques (voir *supra*).

Depuis une dizaine d'années, les artistes visuels se sont emparés des technologies d'IA pour faciliter et améliorer leur processus créatif, expérimenter de nouvelles techniques, s'ouvrir à d'autres styles, réaliser des installations interactives avec le public. Dans tous ces usages, **l'IA agit à la fois comme outil d'aide à la création et catalyseur pour l'inspiration**.

En 2016, *The Next Rembrandt*, un tableau composé par une IA à partir de l'analyse de 346 œuvres du maître hollandais est présenté à Amsterdam.

En 2018, le collectif français *Obvious*, dont la mission d'information a auditionné l'un des membres, réalise une série de onze portraits représentant une famille bourgeoise fictive des XVIII^e et XIX^e siècles, les Belamy, grâce à une IA entraînée à partir de milliers de portraits de la peinture classique. L'un des portraits de cette série, celui d'Edmond de Belamy, est vendu chez Christie's pour près d'un demi-million de dollars, une première pour une œuvre générée par l'IA.

En 2022, l'installation « *Hyperphantasia, des origines de l'image* » de l'artiste plasticienne française Justine Emard, conçue en collaboration avec des archéologues et des préhistoriens, fait appel à l'IA générative : des GAN, entraînés sur un vaste corpus d'images de la grotte Chauvet-Pont d'Arc, ont permis d'obtenir des variantes de ces dessins préhistoriques.

En 2025, l'artiste turco-américain Refik Anadol présente au musée Guggenheim de Bilbao une installation audiovisuelle qui, à l'aide de l'IA, réimagine l'héritage architectural de Franck Gehry. Le logiciel utilisé a été entraîné pendant des mois avec des images, des croquis, des plans du célèbre architecte afin de traduire son vocabulaire en paysage numérique de formes, de couleurs et de mouvements en constante évolution.

b) Dans le cinéma et l'animation

Tout comme les arts visuels, les filières de l'image animée (cinéma, audiovisuel, jeu vidéo) ont depuis une dizaine d'années recours à l'IA pour ses potentialités techniques. La puissance inédite des modèles d'IA générative, comme Midjourney ou ChatGPT, bouleverse toutefois en profondeur la manière dont sont produites les œuvres cinématographiques, audiovisuelles et vidéoludiques.

Une étude publiée en avril 2024 par le Centre national de la cinématographie et de l'image animée (CNC)¹, dans le cadre de son Observatoire de l'IA, et dont les principaux enseignements ont été présentés à la mission d'information, offre une cartographie très intéressante des usages actuels et potentiels de l'IA dans ces filières.

Il en ressort les grands constats suivants :

- les applications de l'IA sont nombreuses (une soixantaine de cas d'usage a été identifiée dans le cadre de cette étude), sur l'ensemble des chaînes de valeur du cinéma, de l'audiovisuel et du jeu vidéo, avec un potentiel plus marqué pour les étapes de post-production, d'animation et d'effets visuels, et pour le secteur des jeux vidéo ;

En phase de préproduction, l'IA est notamment utilisée pour :

- analyser les attentes du public, repérer les éléments narratifs les plus susceptibles de lui plaire et orienter les choix créatifs en fonction des attentes identifiées ;
- générer des scripts à partir de l'analyse de scénarios existants ;
- aider à l'écriture de scénarios ;
- concevoir des *storyboards*.

En phase de production, l'IA permet d'automatiser des tâches logistiques comme l'élaboration des plannings de tournage.

En phase de postproduction, l'IA ouvre de nouvelles opportunités en matière de :

- montage vidéo et son ;
- création d'effets spéciaux (réalisation de visuels plus complexes et plus réalistes, création de scènes de masse sans figurants réels, vieillissement ou rajeunissement d'acteurs, « renaissance » d'acteurs décédés) ;
- composition de bandes originales ou d'effets sonores personnalisés ; techniques de post-synchronisation, doublage, synchronisation labiale et sous-titrage.

- les opportunités offertes par l'IA sont de divers ordres : gain de productivité, augmentation des capacités de production, stimulation de la créativité *via* les échanges homme-machine et la réallocation du temps vers les

¹ « Quel impact de l'IA sur les filières du cinéma, de l'audiovisuel et du jeu vidéo ? », étude menée par le Centre national du cinéma et de l'image animée (CNC) et BearingPoint, avril 2024.

tâches à plus forte valeur créative, ouverture sur de nouveaux possibles techniques, créatifs et économiques, meilleure accessibilité des œuvres ;

- les acteurs de ces filières n'adoptent pas l'IA à la même vitesse, ni dans les mêmes proportions, d'où un impact-métier de l'IA à géométrie variable : les acteurs de grande taille et ceux davantage tournés vers l'innovation sont plus susceptibles de s'adapter aux nouveaux usages que ceux de petite taille et ceux moins directement concernés par la technologie.

c) Dans la musique

Bien que l'attrait du secteur musical pour les technologies informatiques puis numériques remonte à plusieurs décennies, **les capacités désormais offertes par l'IA générative sont sans commune mesure** pour les acteurs de la filière.

L'IA leur permet, entre autres :

- d'améliorer la qualité sonore, en corrigeant les imperfections audio, en ajoutant des effets sonores, en augmentant la dynamique ;
- d'explorer de nouveaux territoires sonores, en créant de nouvelles combinaisons de sons, en générant des sons inhabituels ou en modifiant les sons existants ;
- de générer, à partir de l'analyse de milliers de morceaux et de l'extraction de leurs éléments constitutifs (mélodies, harmonies, rythmes...) de la musique originale dans une multitude de styles et pour diverses applications (musique d'ambiance, musique de film, musique de publicité...).

d) Dans la littérature

En comparaison des autres secteurs, celui de la création littéraire semble pour le moment un peu moins perméable à l'IA, même si ChatGPT est désormais capable de générer des écrits originaux ou des récits à la manière de tel ou tel auteur, disponibles à la vente sur de grandes plateformes en ligne.

Des cas d'usage explicites sont toutefois à mentionner :

- en 2018, une petite maison d'édition française, Jean Boîte, publie un roman en anglais, *1 the Road*, exclusivement écrite par une IA entraînée sur des livres classiques américains ;
- en 2024, l'autrice japonaise Rie Kudan, lauréate du prestigieux prix littéraire Akutagawa pour son livre « *La Tour de compassion de Tokyo* », a révélé qu'environ 5 % de son roman futuriste avait été écrit par ChatGPT, précisant que le logiciel l'avait aidée à libérer son potentiel créatif.

D'autres cas d'usage, plus expérimentaux et confidentiels, ont également sans doute cours.

B. LA CONCURRENCE D'UNE NOUVELLE FORME DE CRÉATION QUI POSE UN DÉFI ÉCONOMIQUE, SOCIAL ET CULTUREL

1. Les contenus culturels sortants : des productions parasites

L'atteinte au droit d'auteur à l'entrée des modèles d'IA (voir *infra*), qui se matérialise par la collecte et la transformation d'une vaste quantité de données d'entraînement à contenus culturels protégés, entraîne **une concurrence déloyale en sortie de modèle, tout aussi grave voire plus inquiétante pour l'avenir de la création artistique humaine.**

Cette concurrence peut même être qualifiée de parasite puisque, dans le cas d'espèce, un agent économique - le fournisseur d'IA - s'immisce dans le sillage d'un autre - le créateur - afin de tirer profit, sans rien dépenser, de ses efforts et de son savoir-faire.

En effet, l'IA générative produit **des données sortantes qui entrent directement en concurrence avec les données-œuvres humaines ayant servi à leur élaboration.** Ces contenus culturels concurrents, la professeure Joëlle Farchy les nomme des « **quasi-œuvres** » dans son rapport remis au CSPLA¹ pour signifier qu'ils **ne remplissent pas la condition d'originalité** propre aux œuvres de l'esprit, lesquelles sont, de fait, protégées par le droit d'auteur.

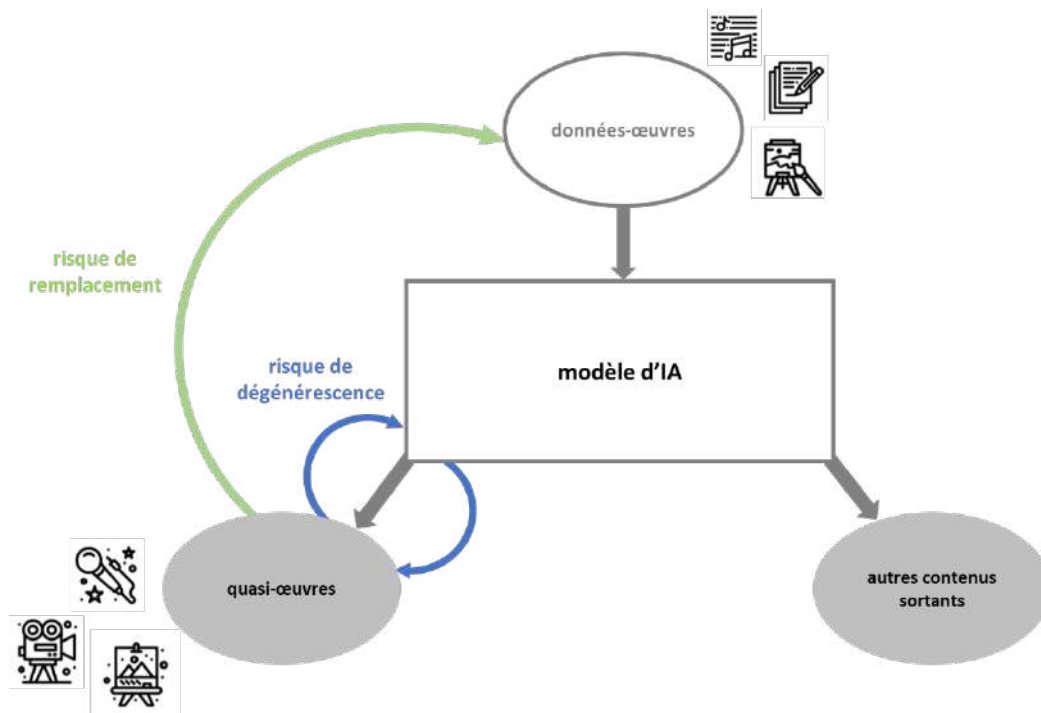
La notion d'originalité d'une œuvre

La notion d'originalité d'une œuvre n'étant pas définie dans la loi, c'est la jurisprudence qui est venue la caractériser et la poser en condition essentielle à la protection par le droit d'auteur.

Selon différents arrêts de la Cour de Cassation, l'originalité s'entend comme « *le reflet de la personnalité de l'auteur* », « *l'expression ou l'empreinte de la personnalité du créateur* » ou « *l'empreinte du talent créateur personnel* ». La notion subjective d'originalité s'oppose ainsi à la notion objective de nouveauté, qui renvoie à l'absence d'antériorité.

En droit européen, la notion d'originalité est définie comme « *la création intellectuelle propre à (son) auteur* » (CJCE, 16 juill. 2009, aff. C-5/08, pt 35).

¹ Voir rapport précité.



Source : « Rémunération des contenus culturels utilisés par les systèmes d'intelligence artificielle », projet de rapport - volet économique, mission confiée par le Conseil supérieur de la propriété littéraire et artistique, Joëlle Farchy et Bastien Blain, mai 2025.

Ainsi que l'explique Joëlle Farchy dans son rapport précité, l'IA engendre, avec ses « quasi-œuvres », un effet d'éviction des œuvres humaines :

- par les prix, du fait de sa capacité à produire plus vite et moins cher que ne le font les hommes ;
- par les quantités, compte des volumes de contenus qu'elle est en mesure de générer, lesquels viendront saturer le marché au détriment des œuvres humaines.

Cette déstabilisation économique du secteur de la création, inédite par sa nature et son ampleur, fait craindre un double risque :

- un risque de substitution des créateurs par les machines (« le grand remplacement ») ;
- un risque de dévitalisation de la création artistique se traduisant par une production standardisée et déshumanisée.

Des prévisions de perte de revenus pour les créateurs encore incertaines

Dans le domaine de la musique, une étude de la Confédération internationale des sociétés d'auteurs et de compositeurs (Cisac), publiée en novembre 2024 prévoit une diminution des revenus des créateurs dans le monde de 24 %, soit 10 milliards de dollars d'ici à 2028, sous l'effet de l'arrivée massive d'œuvres générées par l'IA.

Pour autant, les auditions menées par la mission d'information ont montré que si de plus en plus de contenus musicaux produits par l'IA étaient bien présents sur les plateformes de *streaming* – entre 15 % et 20 % à l'heure actuelle -, ils ne généraient que des revenus marginaux, de l'ordre de 0,5 %. Il est donc à ce stade trop tôt pour extrapoler le lien entre quantité de contenus générés par l'IA et perte de revenus pour les créateurs.

2. Une menace pour les métiers de la création

a) Une menace plus ou moins forte selon les métiers

L'impact de l'IA sur les métiers de la création peut, à ce jour, difficilement faire l'objet d'un état des lieux global et exhaustif, les études sur le sujet restant encore limitées et souvent sectorielles. Néanmoins, à partir de celles disponibles – en particulier l'étude précitée du CNC - et des remontées de terrain des professionnels, quelques grandes tendances se dessinent.

Premier constat, **l'IA générative**, en se substituant à certaines tâches nécessitant des compétences avancées et spécialisées, **concurrence davantage les emplois qualifiés que ceux non qualifiés**. Contrairement aux précédentes révolutions industrielles, ce sont les « cols blancs » qui sont aujourd'hui principalement menacés.

Deuxième constat, **l'IA n'affecte pas ces emplois qualifiés dans les mêmes proportions**. Leur exposition au risque de remplacement dépend du degré de maturité technologique de leur domaine d'activité et du taux de pénétration des applications d'IA au sein de celui-ci. Par exemple, l'IA ayant fait, au cours des dernières années, des avancées particulièrement notables sur la compréhension du texte et la génération de voix, est plus à même d'affecter les métiers de l'écriture et de la voix, que ceux de la vidéo et de la modélisation 3D, secteurs encore en développement technologique.

Troisième constat, **l'IA bouleverse potentiellement plus fortement les métiers de la création à forte composante technique plutôt que ceux à dominante artistique**. Par exemple, des métiers comme réalisateur, compositeur ou comédien ont un potentiel d'automatisation plus faible que des métiers comme monteur son, bruiteur ou artiste en effets visuels.

b) Un risque existentiel pour certaines professions

Si l'impact de l'IA sur les métiers de la création dépend de multiples facteurs et n'est, de ce fait, pas facilement analysable, **certaines professions, dont la mission d'information a tenu à échanger avec les représentants, sont indéniablement touchées de plein fouet**. Pour elles, leur survie est en jeu.

Les graphistes sont particulièrement menacés par des systèmes d'IA générative d'images comme Midjourney. Au Royaume-Uni, la *Society of Authors* a estimé *via* un sondage qu'environ un quart des illustrateurs a déjà perdu du

travail en raison de l'IA générative. En outre, plus d'un tiers des illustrateurs interrogés déclarent que leurs revenus ont diminué en valeur à cause d'elle.

Les doubleurs sont très durement exposés à la concurrence de systèmes d'IA générative comme HeyGen, Eleven Labs ou Deepbub qui permettent de cloner des voix et de traduire des vidéos en plusieurs langues tout en adaptant les mouvements des lèvres. L'utilisation de plus en plus systématique de ces outils se fait au détriment de l'enregistrement en studio des doublages par des comédiens. Selon une étude du Datalab du groupe Audiens, **15 000 emplois directs de comédien de doublage sont ainsi menacés en Île-de-France.**

Les traducteurs sont également fortement fragilisés par le déploiement de systèmes d'IA générative de traduction comme DeepL. Leurs commandes diminuent, elles changent de nature (moins de traductions complètes, davantage de prestations de correction de traductions effectuées par l'IA), entraînant une perte de revenus. Au sein de la profession, certaines spécialités comme la traduction de BD ou de livres audio sont plus affectées que d'autres (comme la traduction littéraire).

c) Un besoin urgent de formation et d'accompagnement des professionnels

La mission d'information estime que les secousses produites par l'IA générative sur les métiers de la création, d'intensités variables et aux conséquences plus ou moins lourdes - allant de la transformation des pratiques professionnelles à la disparition pure et simple de certaines professions - rendent indispensables, à court terme, la mise en place, par les industries culturelles et créatives et les pouvoirs publics, d'actions de formation et d'accompagnement pour les professionnels concernés.

L'objectif, certes délicat mais primordial, est de leur donner les moyens de prendre la vague de l'IA, tout en préservant, quand cela est possible, leur niveau de savoir-faire « traditionnel ».

3. Un risque d'uniformisation et de déshumanisation de la création artistique

Au défi économique et social que pose l'essor de l'IA générative à l'écosystème de la création **fait écho un défi culturel et anthropologique.**

La concentration de la puissance technologique entre quelques grands acteurs dominants, la standardisation des algorithmes et la persistance inévitable de biais dans les bases de données utilisées pour l'entraînement des modèles d'IA font courir un risque **d'uniformisation des contenus culturels générés.**

Sachant que ces « quasi-œuvres » sont elles aussi appelées à entrer dans les prochaines bases d'apprentissage, il y a là une forme de **consanguinité porteuse de danger pour la diversité et la richesse des créations à venir.**

Ce risque d'appauvrissement artistique de la création se doublerait d'un risque de perte d'authenticité et d'originalité, qualités qu'à ce jour, seul un artiste

humain peut apporter grâce à son vécu, sa sensibilité, son savoir-faire et son expérience.

La mission d'information estime qu'il est encore temps de ne pas emprunter cette voie déshumanisée.

III. UN ENCADREMENT JURIDIQUE EN CONSTRUCTION AU NIVEAU MONDIAL

Le développement extrêmement rapide des IA génératives au début de la décennie 2020 et surtout la forte médiatisation de ChatGPT, développé par la société OpenAI qui réalise déjà un chiffre d'affaires annuel de 10 milliards de dollars et compte 500 millions d'utilisateurs par semaine, a révélé le **retard des législations sur la technologie, en particulier en matière de droits d'auteur.**

Dès le 19 septembre 2023, plusieurs associations d'écrivains américains, dont George R.R Martin, auteur de la saga *Games of Thrones*, ont déposé une plainte devant le tribunal fédéral de New York. Les plaignants accusent ChatGPT d'avoir été « entraîné » en utilisant leurs œuvres, sans autorisation ni rémunération, ouvrant la voie à des créations entièrement générées par l'IA « à la façon » de tel ou tel auteur.

Comme on l'a vu, l'essor des modèles de fondation d'IA générative pose de nombreuses questions juridiques, économiques et éthiques, qu'il est d'autant plus difficile de résoudre que l'on ignore encore leurs capacités à court et moyen terme, tant la technologie évolue rapidement et s'étend à un plus grand nombre de domaines.

Dans ce contexte, l'Europe se retrouve entraînée dans une forme de **course contre la montre** pour tenter de réguler l'IA avant qu'il ne soit trop tard, tout en ayant le souci de ne pas brider l'innovation.

Les incertitudes juridiques qui entourent l'IA concernent à la fois **l'amont, avec la question de la rémunération des données, et l'aval, avec le statut juridique des œuvres générées par IA.**

- **En amont**, aucune législation spécifique ne traitait d'IA en Europe avant l'adoption, en 2025, du Règlement sur l'Intelligence Artificielle (RIA). Jusqu'alors, les concepteurs d'IA ont dû évoluer dans un cadre juridique incertain en matière de protection des droits d'auteur, arrêté par la directive *ad hoc* de 2019. Il n'est cependant pas certain que le RIA soit en mesure d'apporter la sécurité juridique nécessaire.

- **En aval**, le droit d'auteur doit composer avec l'afflux (voir partie II du présent rapport) des œuvres générées ou assistées par IA, là encore sans qu'un cadre dédié, protecteur des ayants droit, ait été édicté.

Les prochains mois seront décisifs pour poser les fondations d'une législation adaptée à cette nouvelle technologie.

A. LA DIRECTIVE DU 17 AVRIL 2019 SUR LES DROITS D'AUTEUR : UN SOCLE DÉSORMAIS DÉPASSÉ

La directive européenne du 17 avril 2019 sur les droits d'auteur et les droits voisins dans le marché unique numérique (DAMUN) a posé un cadre à l'époque très avancé et protecteur, qui a su tenir compte des spécificités de l'espace numérique.

Cette législation a cependant été rapidement rattrapée par l'irruption des IA génératives. Elle constitue cependant encore aujourd'hui la base juridique qui a permis leur développement en Europe. **Les fournisseurs d'IA ont donc su utiliser à leur profit le flou juridique d'un texte qui n'avait pas été prévu pour cet usage.**

1. Les exceptions « *text en data mining* » (TDM)

Le titre II de la directive prévoit **deux exceptions** au droit d'auteur, pour la fouille de textes et l'extraction de données, dites TDM (« *text and data mining* »). Elles ont été transposées en droits français aux articles 122-5 et 122-5-3 du code de la propriété intellectuelle.

La fouille de textes et l'extraction de données

Ce mécanisme est défini à l'article 2 de la directive précitée, qui précise qu'il faut entendre par fouille de texte et de données « *toute technique d'analyse automatisée visant à analyser des textes et des données sous une forme numérique afin d'en dégager des informations, ce qui comprend, à titre non exhaustif, des constantes, des tendances et des corrélations* ».

- **La première exception**, prévue à l'article 3 de la directive, est réservée aux reproductions et extractions « effectuées par des organismes de recherche et des institutions du patrimoine culturel, en vue de procéder, à des fins de recherche scientifique, à une fouille de textes et de données sur des œuvres ou autres objets protégés auxquels ils ont accès de manière licite. ». Présente de longue date dans le droit de la propriété intellectuelle, cette exception permet aux laboratoires de recherche publics de disposer en toute sécurité des données présentes dans l'espace numérique à des fins de recherche.

- **La seconde exception**, qui figure à l'article 4 de la directive, prévoit une exception aux droits d'auteur, accessible à tous, « *pour les reproductions et les extractions d'œuvres et d'autres objets protégés accessibles de manière licite aux fins de la fouille de textes et de données* ».

Cette dernière exception TDM était initialement pensée pour **permettre les activités scientifiques menées par des institutions non publiques**. Rejetée dans l'étude d'impact initiale, car considérée comme

attentatoire au droit d'auteur, elle a finalement été réintroduite par voie d'amendement.

a) Les limites de l'exception TDM

L'exception TDM prévue à l'article 4 de la directive précitée de 2019 est cependant limitée de trois manières.

(1) Le respect des contenus protégés

Tout d'abord, ne peuvent faire l'objet de fouilles que les contenus « *accessibles de manière licite* ». Sont donc exclus par définition les contenus **protégés**, par exemple, *via* un accès payant comme l'essentiel des productions de presse ou audiovisuels. Cette précision est d'autant plus essentielle que la très grande majorité des contenus culturels sont désormais accessibles en ligne, que ce soit gratuitement ou moyennant l'acquiescement d'une certaine somme. **La fouille de données ne peut donc trouver à s'appliquer que pour des œuvres librement accessibles.**

(2) La faculté d'« *opt-out* »

Ensuite, le 3^e paragraphe de l'article 4 prévoit que « *L'exception ou la limitation prévue au paragraphe 1 s'applique à condition que l'utilisation des œuvres et autres objets protégés visés audit paragraphe n'ait pas été **expressément réservée par leurs titulaires** de droits de manière appropriée, notamment par des procédés lisibles par machine pour les contenus mis à la disposition du public en ligne* ».

Les titulaires de droit peuvent donc s'opposer de manière **explicite** à l'utilisation de leurs données ; on parle alors de **clause de retrait** ou « *d'opt-out* ». Les formalités ne sont pour autant pas précisées, l'article indiquant simplement que la réserve de droits doit être faite « *de manière appropriée* », à l'aide de procédés lisibles par machine. Ce flou est révélateur du fait qu'en 2019, cette disposition était considérée comme non prioritaire, l'attention étant alors plutôt focalisée sur les contenus protégés et sur le piratage.

Il n'existe donc pas de méthodologie unifiée ou simple pour que les ayants droit puissent faire valoir l'option d'*opt-out*, à tel point que certains ont pu s'interroger devant la mission d'information sur la compatibilité de cette disposition avec l'article 5 de la Convention de Berne pour la protection des œuvres littéraires et artistiques du 9 septembre 1886, qui indique que « *la jouissance et l'exercice de ces droits [les droits d'auteur] ne sont subordonnés à aucune formalité* ».

(3) Le test en trois étapes

Enfin, et comme **toutes** les limitations au droit d'auteur, l'exception demeure soumise au respect du **test dit en trois étapes**. Défini dans plusieurs traités internationaux, ce test, utilisé par les juridictions comme par les législateurs, est repris au paragraphe 5 de l'article 5 de la directive

européenne du 22 mai 2001 sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information.

Le test en trois étapes

L'idée générale est que les exceptions au droit d'auteur doivent être réduites et de portée étroite. Concrètement, le juge ou le législateur sont amenés à poser trois questions :

1. L'exception est-elle bien autorisée ?
2. L'exception porte-t-elle atteinte à l'exploitation normale de l'œuvre ?
3. L'exploitation porte-t-elle un préjudice injustifié aux droits des auteurs ?

2. Une exception détournée ?

Depuis son adoption, l'exception TDM a été interprétée de manière très large, y compris pour des usages non prévus en 2019. De fait, les fournisseurs d'IA ont bénéficié d'un cadre juridique, certes ambigu, mais bien réel, pour justifier leur utilisation des données, et l'exception a été largement mise à profit pour le lancement des IA génératives et sert encore aujourd'hui de fondement juridique à leur entraînement.

La mission d'information, lors de son déplacement à Bruxelles le 12 mai 2025, a ainsi pu échanger avec l'eurodéputé Axel Voss, rapporteur de la directive de 2019 sur les droits d'auteur, qui a rappelé que cette exception n'avait pas été introduite à l'origine pour des modèles d'affaires, mais que sa rédaction peu satisfaisante avait ouvert le champ à un usage plus large. Ainsi, l'étude préparatoire de mars 2014 consacrée à cette exception, réalisée à la demande de la Commission européenne, ne contient **aucune référence à un usage par les machines**, et met au contraire l'accent sur l'intérêt de l'exception en matière de recherche scientifique¹. La professeure Alexandra Bensamoun, lors de son audition devant la commission le 7 mai 2025, a exprimé la même position : *« Rejetée au stade de l'analyse d'impact, car considérée comme trop attentatoire au droit d'auteur et aux droits voisins, elle a finalement été introduite par voie d'amendement à l'article 4 de la directive. On ne voit pas tellement d'indices qui permettraient d'exclure l'entraînement de l'IA générative du champ d'application de la directive ».*

Cependant, au regard des limites à l'exception, **trois incertitudes** demeurent, qui font peser **un réel danger à moyen terme sur l'ensemble des modèles d'IA**. En effet :

- aucune disposition ne permet d'appliquer l'exception à des contenus **protégés**, utilisés sans le consentement des ayants droit. Or il existe

¹<https://op.europa.eu/en/publication-detail/-/publication/074ddf78-01e9-4a1d-9895-65290705e2a5/language-en>

peu de doutes, y compris de la part des fournisseurs, que de tels contenus ont bien été moissonnés pour l'entraînement des IA ;

- si le mécanisme d'*opt-out* n'est pas précisé de manière explicite, il ne fait là encore guère de doute que **les fournisseurs d'IA ont pu passer outre et utiliser des contenus libres d'accès**, mais dont les ayants droit avaient explicitement interdit l'usage ;

- enfin, il est loin d'être certain, compte tenu du développement des IA, que leur utilisation des données respecterait le test en trois étapes qui fonde le droit d'auteur dans le droit international.

En la matière, **les fournisseurs d'IA ont semblé agir suivant une logique propre à la pensée libertarienne américaine**, théorisée en particulier par l'entrepreneur Peter Thiel¹ il y a dix ans. Celle-ci consiste, pour développer une entreprise et acquérir rapidement une position de monopole, à aller vite, à ignorer les règles, et à régler ultérieurement les questions juridiques. Cette philosophie a guidé l'essentiel du développement économique d'Internet depuis le début des années 2000, la législation ayant toujours un temps de retard sur l'innovation et les pouvoirs publics des différents pays étant toujours hésitants à freiner le développement des entreprises de haute technologie.

3. L'intervention des tribunaux

La plupart des entreprises qui développent des IA génératives affirment que les doctrines de l'utilisation équitable, appelée « *fair use* » aux États-Unis, protègent leur pratique consistant à récupérer des données et à les utiliser pour l'entraînement sans compenser – ni même créditer – leurs propriétaires. Elles soutiennent qu'elles ont le droit d'utiliser tout contenu accessible au public pour l'entraînement et que leurs modèles produisent en fait des œuvres « transformatrices » (voir partie IV du présent rapport).

Dans un contexte juridique très mouvant, qui peine à suivre le rythme de la technologie, **les tribunaux des deux côtés de l'Atlantique sont de facto appelés à pallier les insuffisances et les incertitudes des législations existantes.**

a) En Europe : la prudence dans l'utilisation de l'exception TDM

(1) Une première décision à Hambourg...

Le Tribunal de Hambourg a rendu, le 24 septembre 2024, une première décision concernant l'exception TDM pour l'entraînement d'une IA.

Dans le cas d'espèce, l'organisme à but non lucratif allemand LAION, dont l'activité principale est la mise à disposition publique et gratuite d'une

¹ "Zero to One: Notes on Start Ups, or How to Build the Future", Peter Thiel, 2015

banque de données d'entraînement pour les IA, a publié un set de données nommé LAION-5B, utilisé pour entraîner certains systèmes d'IA génératives. Il comprenait en particulier un lien hypertexte conduisant vers une image mise en ligne et vendue sur le site internet du photographe Robert Kneschke. Informé de l'usage fait de sa création, le photographe a accusé LAION d'avoir téléchargé une copie d'une de ses photographies alors que les conditions d'utilisation du site interdisaient l'utilisation des images par des « programmes automatisés ». LAION a invoqué, pour sa défense, l'exception TDM et le tribunal a considéré que les reproductions opérées par LAION avaient bien été effectuées à des **fins scientifiques** , et étaient donc couvertes par l'exception à **des fins de recherches** prévue à l'article 3 de la directive.

Le tribunal a précisé que celle-ci ne doit pas être entendue d'une manière **trop restrictive** . Les juges ont ainsi reconnu que, bien que la création du jeu de données ne puisse pas être en tant que telle considérée comme un gain de connaissance, cette base de données en constituait une étape essentielle. En ce sens, le tribunal a considéré que le fait que le jeu de données soit publié gratuitement et ouvertement, donc sans objectif commercial, pouvait permettre à des chercheurs d'en bénéficier et donc de créer un gain de connaissance.

Dans une autre mesure, cette décision a également précisé l'exception TDM posée à l'article 4 , plus particulièrement en ce qui concerne la notion d'*opt-out*. La DAMUN indique que cette réserve peut être exprimée notamment dans le cadre de conditions générales d'utilisation mais « *au moyen de procédés lisibles par machine* ». En l'espèce, l'*opt-out* était inscrit dans les **conditions générales** du site internet sur lequel la photographie était reproduite, sous la forme d'un **texte** et non d'un **procédé technique particulier** , ce que les juges ont trouvé **suffisant** pour que l'*opt-out* soit considéré comme valablement manifesté. **L'exception de l'article 4 ne pouvait donc pas s'appliquer en l'espèce car l'opt-out avait été suffisamment exprimé.**

La décision du tribunal allemand, finalement favorable au développement de l'IA, n'a donc été justifiée que par le **statut à but non lucratif de LAION** , et par le fait que son action pouvait participer d'un effort de recherche scientifique.

(2) ...en attendant la Cour de Justice de l'Union européenne

La Cour de Justice de l'Union européenne (CJUE) devra très vraisemblablement se prononcer, à terme, sur l'étendue de l'exception TDM. Elle pourrait en avoir l'occasion en 2026. En effet, à l'initiative de l'éditeur hongrois *Like Company*, une question préjudicielle lui a été adressée le 3 avril 2025. L'éditeur conteste l'utilisation par le moteur de recherche Google et son IA Gemini de contenus de presse, ouvrant la voie à une jurisprudence globale sur l'exception TDM.

b) Aux États-Unis : quelles limites au « fair use » ?

(1) Un « fair use », très largement invoqué

Les fournisseurs américains d'IA ont largement utilisé les données disponibles en ligne pour entraîner leurs IA en s'abritant derrière la notion anglo-saxonne de « fair use », ou « usage loyal ». Elle correspond en partie aux exceptions prévues dans le droit français et européen.

Le fair use aux États-Unis

Le « fair use » est défini à l'article 107 du Code des États-Unis. Son usage a donné lieu à une abondante jurisprudence. Il vise principalement à permettre l'utilisation de contenus protégés dans un objectif de copie privée, de recherche, de satire, ou encore d'enseignement. Le juge l'apprécie au regard de quatre critères :

- (1) *L'objectif et la nature de l'usage, notamment s'il est de nature commerciale ou éducative et sans but lucratif ;*
- (2) *la nature de l'œuvre protégée ;*
- (3) *la quantité et l'importance de la partie utilisée en rapport à l'ensemble de l'œuvre protégée ;*
- (4) *les conséquences de cet usage sur le marché potentiel ou sur la valeur de l'œuvre protégée.*

Une conception très large de la doctrine du « fair use » a été mise en avant par les concepteurs des modèles d'IA pour justifier leur usage des données, suivant en cela une logique que le monde numérique applique depuis les origines. Ainsi, Mustafa Suleyman, responsable de l'Intelligence artificielle de Microsoft, actionnaire de la société OpenAI, a publiquement déclaré sur la chaîne américaine CNBC¹ en 2024 : « *Je pense qu'en ce qui concerne les contenus qui sont déjà présents sur le web ouvert, le contrat social depuis les années 90 a été qu'on peut les utiliser de manière équitable. N'importe qui peut les copier, les recréer, les reproduire. C'est le libre usage, si vous voulez, c'est ce que l'on a compris* »

(2) Une première décision dans le Delaware

Par un jugement du 11 février 2025, la Cour fédérale du Delaware a cependant reconnu une **violation du droit d'auteur** par la société *Ross Intelligence* pour l'utilisation non autorisée d'éléments de résumés de la base de données juridiques « *Westlaw* » appartenant à Thomson Reuters, dans le but d'entraîner son IA juridique, laquelle est destinée à analyser des documents et à effectuer des recherches.

La société Ross estimait que son utilisation des résumés protégés par le droit d'auteur était légalement défendable car elle était « transformatrice », c'est-à-dire qu'elle réutilisait les résumés pour servir une fonction ou un

¹ <https://www.theverge.com/2024/6/28/24188391/microsoft-ai-suleyman-social-contract-freeware>

marché **sensiblement différent**. Le tribunal n'a cependant pas retenu ce raisonnement qui repose sur le « *fair use* », considérant que l'usage n'était pas suffisamment transformateur et servait le même objectif commercial que Thomson Reuters.

Ce jugement, qui devrait faire l'objet d'un appel, ne porte cependant pas un coup d'arrêt à l'usage des données. En effet, **la décision du tribunal établit une distinction entre l'IA générative et l'IA spécialisée qu'utilisait Ross**. Il n'est pas donc pas évident que cette logique puisse s'appliquer aux IA génératives.

Lors de son audition devant la commission, la professeure Alexandra Bensamoun a évoqué cette affaire, en établissant un parallèle avec le droit européen : « *Il y a eu récemment une décision intéressante dans une affaire Ross Intelligence , même si elle ne portait pas précisément sur l'IA. Le juge a estimé que l'opération en cause ne relevait pas du fair use. Cela signifie que le droit exclusif n'a pas été respecté. Or, aux États-Unis, les dommages-intérêts peuvent être punitifs. Ce n'est pas notre système. Nous avons, en principe, une réparation intégrale : rien que le préjudice, mais tout le préjudice. En propriété intellectuelle, ce principe a été légèrement infléchi, sous impulsion européenne, mais en cas de faute lucrative, notre modèle atteint ses limites. Aux États-Unis, les dommages-intérêts excèdent largement le préjudice subi et vont jusqu'au punitif. C'est pourquoi les contentieux peuvent avoir des conséquences considérables sur le marché. Aussi, il est probable que certaines procédures s'éteignent par transaction. C'est une culture que nous gagnerions à acquérir en France.* »

(3) Une forte pression de la nouvelle administration américaine pour protéger les géants de la tech

Un débat extrêmement sensible a actuellement lieu aux États-Unis sur le fondement juridique du « *fair use* » à propos de l'utilisation de données protégées lors de la phase d'entraînement des IA.

Dans un rapport remis en mai 2025, le Bureau du Copyright des États-Unis (*United States Copyright Office*) **émet des doutes sur la légitimité d'un usage massif du *fair use*** pour justifier l'utilisation des données par les concepteurs d'IA¹. Tout en reconnaissant la flexibilité et l'adaptabilité de cette doctrine, le Bureau formule de sérieuses réserves sur son extension à des usages commerciaux : « *Faire un usage commercial d'une vaste collection de contenus protégés pour produire des contenus qui entrent en compétition avec eux sur des marchés existants, tout particulièrement lorsqu'ils ont été récupérés de manière illégale, va au-delà des limites du *fair use**² ». Le rapport plaide pour le développement des accords de licence, sans appeler à une évolution de la législation.

¹ <https://www.copyright.gov/ai/>

² « [...] making commercial use of vast troves of copyrighted works to produce expressive content that competes with them in existing markets, especially where this is accomplished through illegal access, goes beyond established fair use boundaries. »

La réaction du milieu de la tech et des autorités américaines laisse cependant peu de place au doute quant à la position future de l'administration. Ainsi Matt Schruers, président du principal groupement d'intérêt des entreprises de technologie (*Computer and Communications Industry Association* (CCIA)) a déclaré, dans un communiqué en date du 12 mai 2025, que « *le rapport du Copyright Office se met à plusieurs reprises en travers du chemin vers l'intérêt général* », estimant que « *la loi américaine sur le droit d'auteur est suffisante pour répondre aux questions actuelles concernant l'IA et le droit d'auteur* ». Quelques jours après la publication du rapport, **la Maison Blanche licenciait Shira Perlmutter, qui était à la tête de l'agence depuis octobre 2020**, marquant ainsi son opposition à la position du Bureau.

Il reviendra donc aux tribunaux américains, vraisemblablement à la suite de débats juridiques qui dureront des années, de se prononcer définitivement sur l'étendue de la doctrine du « *fair use* » dans le contexte de l'IA, **alors que la pression politique s'exerce très fortement pour une interprétation en ligne avec les *desiderata* du secteur de la tech.**

B. UNE LÉGISLATION EUROPÉENNE QUI CHERCHE UNE DIRECTION

1. La nécessité d'un cadre, la difficulté d'y parvenir

Dès 2021, l'Europe a ressenti le besoin de se doter d'une législation spécifiquement dédiée à l'IA qui prendrait enfin en compte la singularité de cette technologie.

Déposée le 22 avril 2021 par Thierry Breton, alors commissaire européen, la proposition de Règlement sur l'Intelligence Artificielle (RIA) est composée de 84 articles et pose les bases d'une législation globale qui va bien au-delà **des questions de droit d'auteur et de création, lesquelles n'étaient d'ailleurs même pas mentionnées dans la première version**, très favorable au développement d'un secteur européen de l'IA.

De fait, **la question de la protection des droits d'auteur est apparue tardivement dans la discussion, à l'initiative du Parlement européen.** Le trilogue qui s'est tenu entre la Commission européenne, le Conseil et le Parlement européen, du 7 au 12 décembre 2023, a été l'objet d'échanges intenses sur ce sujet et marqué par des actions de lobbying très poussées tant de la part des représentants des ayants droit que de la part de ceux du secteur de la tech.

Ces longues négociations ont souligné de **fortes divergences de vues** entre deux camps, divergences qui structurent encore l'essentiel du débat comme a pu le constater la mission d'information, en particulier lors de son déplacement à Bruxelles :

- **d'un côté**, les promoteurs de l'IA qui estiment qu'une régulation précoce ne permettrait pas à l'Europe de rattraper un retard déjà conséquent

sur les États-Unis et la Chine en entravant toute possibilité de développement d'IA « souveraines » ;

- **de l'autre**, une coalition formée de « techno pessimistes » qui s'inquiètent des dérives possibles de l'IA sur les libertés individuelles, l'emploi et, de manière générale, le tissu social, ou encore les défenseurs des ayants droit, qui jugent que la création de l'IA s'apparente au pillage de contenus couverts par le droit d'auteur.

Dans les débats européens, la France a semblé hésiter entre ces deux pôles, avec toutefois un penchant pour une législation « respectueuse » de l'innovation. Notre pays, avec la société Mistral AI, abrite en effet une entreprise qui semble particulièrement prometteuse, avec une levée de fonds record de 385 millions d'euros et une valorisation qui dépasserait les deux milliards d'euros.

Lors d'un colloque à Londres le 2 novembre 2023, Bruno Le Maire, alors ministre de l'économie, des finances et de la souveraineté industrielle a ainsi semblé pencher en faveur d'une régulation minimum : « *si l'Union européenne veut rester dans la course de l'intelligence artificielle au XXI^e siècle, tous les pays européens doivent mettre en commun leurs forces, leurs compétences, leurs technologies, et investir plus largement et plus rapidement [...] Avant de mettre des obstacles, nous devons donner une impulsion* ». Il a été rejoint par le Président de la République qui, à l'occasion d'un colloque à Paris le 17 novembre 2023, a indiqué qu'il souhaitait une régulation européenne « *maîtrisée et non punitive pour préserver l'innovation* », appelant à « *réguler les usages, davantage que les technologies en tant que telles* ».

La France s'est ainsi souvent trouvée aux côtés de l'Allemagne et de l'Italie pour *a minima* souligner les risques d'une régulation trop ambitieuse. La ligne de fracture a d'ailleurs traversé le gouvernement français, en semblant opposer, selon plusieurs déclarations publiques, les ministères de l'économie et de la culture. Il a fallu attendre le 24 novembre 2023, soit quelques jours avant le trilogue européen, pour que ces deux ministères annoncent travailler à une position commune.

Le 20 décembre 2023, lors d'une table ronde organisée par la commission de la culture¹, le secteur culturel a pu faire valoir publiquement, en des termes parfois véhéments, son incompréhension des positions françaises. Ainsi, le directeur général de la Société des auteurs et compositeurs dramatiques (SACD), Pascal Rogard, déclarait : « *pour la première fois, la France, [...] ne s'est pas positionnée en soutien de la culture ; elle a, au contraire, travaillé à Bruxelles à mettre en place des coalitions avec des pays adversaires du droit d'auteur, afin de constituer une minorité de blocage. Nous sommes donc très déçus. Nous avons entendu un discours qui opposait innovation et création, modernité et droit d'auteur ; or le droit d'auteur, qui est une invention française, n'a jamais bloqué l'innovation !* ». David El Sayegh, directeur général adjoint de la Société des

¹ <https://www.senat.fr/compte-rendu-commissions/20231218/cult.html#toc2>

auteurs, compositeurs et éditeurs de musique (Sacem) soulignait pour sa part « *Je suis moi aussi consterné que la France organise ainsi un système qui favorise l'opacité, alors que nous nous sommes battus avec le Gouvernement pour une protection robuste du droit d'auteur dans le numérique.*

Un léger infléchissement avait cependant été publiquement acté par une déclaration de Jean-Noël Barrot, alors ministre délégué chargé de la transition numérique et des communications à un colloque organisé par la Cnil le 29 novembre 2023 : « *On ne voit pas très bien sur quoi l'intelligence artificielle s'entraînerait si elle venait à décourager les artistes de faire leur travail* ».

Enfin adoptée par le Parlement et le Conseil, le Règlement 2024/1689 du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle (RIA) porte les traces du compromis douloureux entre ces différentes tendances.

2. Les avancées et les incertitudes du RIA

Deux paragraphes du RIA s'intéressent spécifiquement à la préservation des droits d'auteur. Ils concernent deux aspects en apparence distincts, mais en réalité inséparables : **la conformité** au droit de l'Union et le **modèle de résumé**. Faute de clarté suffisante des termes du Règlement, le débat s'est depuis déplacé sur ses **modalités d'application** qui s'avèrent complexes et sont source d'affrontements entre différents intérêts concurrents.

a) La conformité au droit de l'Union

(1) Un simple rappel des règles...

L'article 53, 1, c) du RIA invite les fournisseurs de modèles d'IA qui exercent dans l'Union européenne - ce qui inclut notamment les IA américaines et chinoises - à mettre en place « *une politique visant à se conformer au droit de l'Union européenne en matière de droit d'auteur et droits voisins, et notamment à identifier et à respecter, y compris au moyen de technologies de pointe, une réserve des droits exprimée conformément à l'article 4, paragraphe 3, de la directive 2019/790* ». Si des doutes ont pu exister quant à l'applicabilité de l'exception TDM à l'entraînement des IA, le RIA a donc pour l'instant clos le débat, en s'appuyant **explicitement** sur l'article 4 de la DAMUN, dont l'article 53 constitue d'une certaine manière un « mode d'emploi ».

Les modèles d'IA commercialisés dans l'Union doivent donc respecter **l'ensemble de la législation communautaire protectrice des droits**.

Le paragraphe visé affirme ainsi le principe général du respect de l'acquis communautaire, et prend soin de mentionner explicitement la réserve des droits.

Le considérant 105 du Règlement rappelle ainsi que si l'exception TDM peut bien être utilisée, « *toute utilisation d'un contenu protégé par le droit d'auteur nécessite l'autorisation du titulaire de droits concerné* ».

Le considérant 106 précise la portée de cette obligation : « *Les fournisseurs qui mettent des modèles d'IA à usage général sur le marché de l'Union devraient veiller au respect des obligations pertinentes prévues par le présent règlement. À cette fin, les fournisseurs de modèles d'IA à usage général devraient mettre en place une politique visant à respecter la législation de l'Union sur le droit d'auteur et les droits voisins, en particulier pour identifier et respecter la réservation de droits exprimée par les titulaires de droits conformément à l'article 4, paragraphe 3, de la directive (UE) 2019/790.* »

Cette disposition établit clairement la faculté **d'utiliser l'exception TDM dans le cadre de l'entraînement des IA, ce qui n'était pas acquis avant son adoption tant le développement de ces technologies semblait s'éloigner de la DAMUN.**

Le RIA comporte cependant une ambiguïté. En effet, la DAMUN établit **deux ensembles distincts** (voir *supra*) : les contenus **libres d'accès**, pour lesquels les ayants droit doivent manifester leur faculté d'*opt-out*, et les contenus **protégés**, pour lesquels le droit d'auteur s'applique et où un accord doit être recherché quel que soit l'usage. **Le RIA ne donne pas plus de détails que la DAMUN tant sur les modalités de réserve des droits que sur la protection des contenus protégés.**

À bien des égards, et pris isolément, l'article 53,1,c) du RIA se limite donc à un **rappel** des règles contenues à l'article 4 de la DAMUN.

(2) ... qui doit être complété par un code de bonne pratique

Il existe donc **une très forte incertitude sur les obligations normatives qui pèsent réellement sur les fournisseurs d'IA**, cette situation étant dommageable aussi bien pour eux que pour les ayants droit.

Le 4. de l'article 53 offre cependant la perspective d'un cadre plus homogène et prescriptif pour assurer le respect des droits. Il prévoit ainsi que « *Les fournisseurs de modèles d'IA à usage général peuvent s'appuyer sur des codes de bonne pratique au sens de l'article 56 pour démontrer qu'ils respectent les obligations énoncées au paragraphe 1 du présent article* ». Cet article 56 du RIA prévoit en effet la publication de tels codes qui couvrent en particulier le domaine de l'article 53.

Le 4. entretient une forme **d'ambiguïté** entre le code et « *une norme harmonisée* » qui a vocation à le remplacer à terme. En effet, il est précisé que : « *Le respect des normes européennes harmonisées confère au fournisseur une présomption de conformité dans la mesure où lesdites normes couvrent ces obligations.* » On peut donc en déduire que, le jour où ces normes seront édictées, leur respect vaudra présomption de conformité et que les codes possèdent une valeur inférieure. En tout état de cause, et en l'absence de

norme, **l'adhésion au code de bonne pratique ne vaut pas en lui-même libération de ses obligations par le fournisseur**, et ce d'autant moins que sa signature demeure facultative. Si les fournisseurs choisissent de ne pas y adhérer, ils doivent donc apporter la preuve « *qu'ils disposent d'autres moyens appropriés de mise en conformité et les soumettent à l'appréciation de la Commission* »

Toute l'attention des fournisseurs et des ayants droit s'est donc déplacée sur le contenu du code de bonne pratique. Conformément à l'article 56 du RIA, il doit traiter de plusieurs problématiques liées au déploiement de l'IA en Europe. La section consacrée au respect des droits d'auteur a cependant monopolisé l'attention médiatique et les différentes actions de lobbying, que ce soit de la part des entreprises du secteur de la tech ou des ayants droit.

- (3) Des versions successives de moins en moins protectrices des droits d'auteur

Les travaux autour de l'élaboration de ce code ont débuté dès l'été 2024, avec une large association des experts et des parties prenantes. Une première version a été rendue publique le 14 novembre 2024, suivie d'une seconde le 19 décembre et d'une troisième le 11 mars 2025.

Le sentiment généralement exprimé par les ayants droit est celui d'une grande déception, les différentes versions du code allant selon eux dans une direction « moins-disante » en termes de régulation, et donc trop favorable aux entreprises technologiques. **D'une certaine manière, les ambiguïtés autour de la directive sur les droits d'auteur, qui n'ont pas plus été tranchés avec l'article 53 du RIA, se concentrent dorénavant sur ce code pourtant non contraignant**, mais où chacun veut y déceler une orientation politique plus ou moins en faveur des ayants droit ou des entreprises technologiques.

L'examen des trois versions successives souligne cependant **un biais certain et de plus en plus accentué en faveur du secteur de la tech**, accusé, comme souvent lors des débats européens, d'avoir exercé d'intenses actions d'influence.

Le tableau suivant permet de comparer l'évolution des termes employés sur trois aspects significatifs :

- (1) lorsque l'opération de moissonnage est réalisée directement par le signataire ;
- (2) le respect de l'accès aux seuls contenus légaux et de la réserve des droits ;
- (3) le contrôle du respect par le concepteur de modèle d'IA de la conformité de la base de données.

	Première version (14 novembre 2024)	Deuxième version (19 décembre 2024)	Troisième version (11 mars 2025)
Accès aux seuls contenus légaux directement moissonnés par les signataires	Les signataires « s'engagent à s'assurer qu'elles disposent d'un accès légal aux contenus protégés et à identifier et se conformer aux réserves des droits exprimées en application de l'article 4 de la DAMUN ¹ »	Les signataires « s'engagent à faire des efforts raisonnables et proportionnés pour s'assurer qu'elles disposent d'un accès légal aux contenus protégés ² »	Les signataires - ne doivent pas « contourner les mesures technologiques [...] conçues pour éviter ou restreindre l'accès aux contenus protégés ³ » - doivent « faire des efforts raisonnables pour exclure de leurs robots d'indexation les sites qui rendent disponibles les contenus protégés [...] et n'ont pas de droits légitimes substantiels sur les contenus (« sites pirates ») ⁴ »
Prise en compte de la réserve des droits pour les contenus directement moissonnés par l'entreprise	(idem) Les signataires « s'engagent à s'assurer qu'elles disposent d'un accès légal aux contenus protégés et à identifier et se conformer aux réserves des droits exprimées en application de l'article 4 de la DAMUN »	Les signataires « s'engagent, au minimum, à identifier et à se conformer, au moyen des technologies les plus avancées, à la réservation des droits ⁵ » exprimées par le protocole robots.txt	Les signataires : - « utilise des robots d'indexation qui suivent le protocole robots.txt » - « font les meilleurs efforts pour identifier et se conformer aux autres protocoles appropriés pour faire respecter la réservation des droits »

¹ “[...] they commit to ensure that they have lawful access to copyright-protected content and to identify and comply with rights reservations expressed pursuant to Article 4(3) of Directive (EU) 2019/790”

² “[...] they commit to making reasonable and proportionate efforts to ensure that they have lawful access to copyright-protected content in accordance with Article 4(1) of Directive (EU) 2019/790.

³ “not circumvent effective technological measures [...] designed to prevent or restrict access to works and other protected subject matter...”

⁴ “make reasonable efforts to exclude from their web-crawling Internet domains that make available to the public copyright-infringing content on a commercial scale and have no substantial legitimate uses (“piracy domains”)”

⁵ “they commit, as a minimum measure to identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790, to employing web-crawlers that read and follow instructions expressed in accordance with the Robot Exclusion Protocol (robots.txt),

Contenus obtenus d'un fournisseur	Les signataires conduisent « <i>des diligences raisonnables en matière de droit d'auteur avant de contracter avec une tierce partie</i> ¹ »	Les signataires doivent « <i>faire des efforts raisonnables et proportionnés pour obtenir l'assurance</i> ² » des fournisseurs.	Les signataires « <i>feront des efforts raisonnables pour obtenir des informations adéquates [...]</i> » sur le respect par le fournisseur de données de la réserve des droits exprimée à l'aide de Robot.txt ³
-----------------------------------	--	--	--

Les trois versions comportent une mention invitant au développement d'un protocole **unifié et standardisé** en lien avec les ayants droit pour faire respecter la réserve des droits et la protection des contenus.

À partir de la deuxième version, les obligations peuvent être modulées en fonction de leur taille et de leurs capacités, « *dans l'intérêt des startups* ».

Force est de constater que les itérations successives du code ne vont pas dans le sens d'une plus grande responsabilité des entreprises technologiques.

(4) La forte opposition du secteur culturel

La parution de la troisième version, le 11 mars 2025, a suscité un très fort rejet des ayants droit, publiquement exprimé sous la forme d'une déclaration conjointe signée par 37 organisations représentatives françaises et européennes⁴. Ce texte, largement diffusé, estime que la troisième version « *interprète de manière erronée le droit d'auteur européen et affaiblit les obligations établies par le règlement sur l'IA lui-même.* »

Les signataires reprochent en particulier à cette version de se contenter d'exiger des fournisseurs d'IA « *des efforts raisonnables* », en particulier lors du recours à des bases de données tierces. En ce qui concerne la réservation des droits, le code « *suggère même que combiner l'adoption de mesures techniques de contrôle d'accès et des « efforts raisonnables » pour exclure une liste limitée de sites de piratage suffiraient à garantir la condition « d'accès licite »*

¹ "Signatories will undertake a reasonable copyright due diligence before entering into a contract with a third party about the use of data sets"

² "Signatories commit to making reasonable and proportionate efforts to obtain assurances from a third party about its compliance with Union law on copyright"

³ "they will make reasonable efforts to obtain adequate information (e.g., by checking the information available on the website of the third parties or requesting information) as to whether works and other protected subject matter that have been scraped or crawled from the internet were collected by employing web-crawlers that read and follow instructions expressed in accordance with the Robot Exclusion Protocol (robots.txt), [...] and any subsequent version of this IETF standard."

⁴ <https://snepmusique.com/actualites-du-snep/declaration-conjointe-des-titulaires-de-droit-sur-le-3eme-projet-de-code-de-bonnes-pratiques/>

aux contenus protégés. » La déclaration estime donc que les entreprises d'IA ne seraient pas soumises à des obligations de transparence sur les méthodes employées pour appliquer la réserve des droits, le protocole robots.txt étant jugé peu efficace.

La mission d'information a pu constater, à l'occasion de ses auditions, le très large consensus des ayants droit sur le manque d'ambition de ce texte voire le recul qu'il entérine, dont certains préféreraient d'ailleurs qu'il ne mentionne pas du tout les droits d'auteur - ce qui ne serait cependant pas compatible avec l'article 56 du RIA -.

Lors de son déplacement à Bruxelles, la mission d'information a échangé avec les personnels de la Commission européenne en charge des différentes itérations du code et a **fermement relayé les préoccupations du secteur**, qui étaient au demeurant déjà connues et analysées.

Le point de vue de la Commission européenne, qui travaille encore sur une nouvelle version du code, **est que ce document ne pourra pas résoudre à lui-seul l'ensemble des problématiques posées par l'IA**, mais plus modestement contribuer à élaborer un cadre juridique jusqu'alors inexistant et surtout à créer les conditions d'un marché de la donnée. Tout en étant consciente des difficultés des ayants droit, la Commission européenne s'efforce de mettre en balance les intérêts, encore considérés comme divergents, des créateurs et des entreprises technologiques, avec l'ambition de voir émerger en Europe une grande industrie du numérique, ce qui passe selon elle par un allègement des contraintes.

Alors que le code de bonne pratique devait être disponible au printemps 2025, et que le dernier alinéa de l'article 56 du RIA fixe comme date limite le **2 août 2025, une quatrième version du code est attendue**, sans qu'une date précise ait pu être fixée. Faute de l'existence de ce code, la Commission devra prendre des actes d'exécution, ce qui ôterait la faculté aux parties prenantes d'approuver explicitement le code, et compliquerait donc la mission de contrôle du respect de la conformité.

b) Le résumé des sources utilisées

(1) Des termes volontairement ambigus

Le d) du 1 du même article 53 impose une **seconde obligation** aux fournisseurs de modèles d'IA : **mettre à disposition du public « un résumé suffisamment détaillé du contenu utilisé pour entraîner les modèles d'IA à usage général, conformément à un modèle fourni par le Bureau de l'IA ».**

Le choix des termes retenus dans le RIA porte encore la trace des oppositions qui se sont manifestées lors de son adoption et qui ont permis d'aboutir à un compromis. Le sens à donner à l'expression « **résumé suffisamment détaillé** » est en effet peu clair : à partir de quel degré peut-on considérer qu'un résumé, qui par nature ne comporte pas l'ensemble des

informations, est « *suffisamment détaillé* » ? À quel niveau d'exhaustivité doit-il correspondre ?

(2) Une obligation indissociable de la conformité

Toute la question, à ce jour encore en suspens, est donc la mise en œuvre concrète de ce modèle de résumé. Le considérant 107 du RIA donne des indications sur sa finalité, sans pour autant clarifier le débat :

« Tout en tenant dûment compte de la nécessité de protéger les secrets d'affaires et les informations commerciales confidentielles, ce résumé devrait être généralement complet en termes de contenu plutôt que détaillé sur le plan technique afin d'aider les parties ayant des intérêts légitimes, y compris les titulaires de droits d'auteur, à exercer et à faire respecter les droits que leur confère la législation de l'Union, par exemple en énumérant les principaux jeux ou collections de données utilisés pour entraîner le modèle, tels que les archives de données ou bases de données publiques ou privées de grande ampleur, et en fournissant un texte explicatif sur les autres sources de données utilisées. Il convient que le Bureau de l'IA fournisse un modèle de résumé, qui devrait être simple et utile et permettre au fournisseur de fournir le résumé requis sous forme descriptive. »

Le considérant assigne donc comme principale finalité au modèle de résumé **d'aider les partis à exercer et faire respecter leurs droits**. Il est donc en lien **direct** avec la politique générale de conformité telle qu'elle ressort de l'article 53, 1, c), à tel point que ces deux aspects sont souvent traités en parallèle.

Si les deux dispositions font l'objet de deux paragraphes distincts, ils doivent donc être lus de manière **complémentaire**. Comme l'indique la professeure Alexandra Bensamoun dans son rapport pour le CSLPA sur la mise en œuvre du RIA¹ de décembre 2024, « *la politique de conformité exigée à l'article 53, 1, c du RIA et la mise à la disposition du public d'un résumé suffisamment détaillé imposé à l'article 53, 1, d sont indissociables. La politique de conformité est le négatif du résumé détaillé : ce que le second dit en plein, la première le dit en creux. Ils composent ainsi les deux faces d'une même obligation : l'obligation de transparence. Le modèle de résumé doit en conséquence intégrer les éléments pertinents de la politique de conformité et notamment ceux relatifs à la clause de réserve de droits* ».

(3) Une mise en œuvre qui interroge

Une fois établie que ce résumé doit être suffisamment détaillé pour permettre aux parties prenantes, en particulier les auteurs, de faire valoir leurs droits, la question de sa mise en œuvre concrète se pose.

¹ <https://www.culture.gouv.fr/fr/nous-connaître/organisation-du-ministère/Conseil-supérieur-de-la-propriété-littéraire-et-artistique-CSPLA/travaux-et-publications-du-cspla/missions-du-cspla/ia-et-transparence-des-données-d-entraînement-publication-du-rapport-d-alexandra-bensamoun-sur-la-mise-en-œuvre-du-règlement-européen-etablissant>

Le résumé doit en effet concilier dès sa conception :

- le **droit légitime des ayants droit** à savoir si leur production a fait l'objet d'une utilisation durant la phase d'apprentissage qui pourrait nécessiter un accord et ouvrir la possibilité d'une compensation financière ;
- et le **secret des affaires** qui, dans le cas d'une IA, repose à la fois sur des algorithmes et sur la sélection des données d'entrée.

Les conditions de respect de ces deux objectifs sont au cœur des débats actuels sur la conception de ce résumé.

La mission d'information a pu mesurer durant ses auditions la difficulté à établir un canevas accepté par tous.

D'un côté, les ayants droit recherchent une forme **d'exhaustivité** cohérente avec le monopole d'exploitation dont ils disposent sur les œuvres protégées.

De l'autre, les entreprises technologiques souhaitent préserver le plus possible la nature des données qu'elles ont été amenées à utiliser durant la phase d'entraînement de l'IA, considérant que cela relève du secret des affaires.

D'une certaine manière, les seconds retiennent le terme de « résumé » au sens premier, quand les seconds s'attachent aux moyens de faire respecter les droits prévus dans le considérant 107.

Pour reprendre une analogie culinaire proposée par la professeure Alexandra Bensamoun dans son rapport précité de décembre 2024, et souvent rappelée durant les auditions de la mission, les données utilisées correspondraient aux **ingrédients** d'un plat, qui peuvent être communiqués sans dommage, les techniques utilisées pour les traiter et les intégrer au modèle constitueraient à la **recette**, qui relève incontestablement du secret des affaires. Le rapport plaide alors pour cette distinction qui permet de préserver l'objectif de la législation européenne, à savoir la **transparence** des données utilisées.

Cette interprétation, que la mission d'information partage sans réserve, est la seule de nature à donner un effet utile aux dispositions du RIA et à respecter l'esprit comme la lettre du considérant 107. Elle n'est cependant pas acceptée à ce stade par les fournisseurs d'IA, qui estiment que la connaissance des données utilisées apporte en elle-même une information capitale à leurs concurrents.

Le secret des affaires

L'application aux données du secret des affaires est un sujet qui n'est pas encore pleinement tranché. La professeure Alexandra Bensamoun, dans son rapport précité pour le CSLPA, propose une analyse détaillée de l'état actuel du droit :

« [...] il faut rappeler que l'invocation du secret des affaires a bien entendu des limites. En droit interne, l'article L. 151-7 du code de commerce dispose que le secret des affaires ne peut être opposé aux autorités juridictionnelles et administratives agissant, notamment, dans l'exercice de leurs pouvoirs d'enquête, de contrôle, d'autorisation ou de sanction. Et l'on notera avec intérêt que, dans l'affaire *Dun & Bradstreet Austria GmbH C-203/22* relative au traitement de données personnelles par une IA, ayant conduit à refuser la conclusion ou la prolongation d'un contrat de téléphonie mobile au motif que la personne ne présentait pas une solvabilité financière suffisante, l'avocat général Jean Richard de la Tour a considéré, le 12 septembre 2024, que le secret des affaires ne pouvait conduire à écarter le droit qu'un individu tire du RGPD de comprendre comment une décision qui l'affecte est prise. Cette position paraît transposable aux droits qu'une personne tient des dispositions de droit d'auteur issues des textes européens. Le secret des affaires ne peut conduire, en vidant de toute substance le résumé suffisamment détaillé, à écarter le droit qu'un titulaire de droits tire du RIA à disposer d'éléments pouvant l'aider « à exercer et à faire respecter les droits que leur confère la législation de l'Union » 58. Enfin, la directive relative au secret des affaires envisage même l'hypothèse d'une règle de l'Union qui exigerait la révélation d'informations au public, y compris des secrets des affaires, pour des motifs d'intérêt public».

Ce point soulève la difficulté, plusieurs fois rappelée durant les auditions, du caractère **public** du résumé, qui doit donc être accessible à tout un chacun, **alors même que sa finalité est plus étroite**, puisqu'il a pour objet essentiel de garantir le respect des prérogatives des ayants droit.

Il est cependant possible d'établir une forme de hiérarchisation entre les données utilisées et donc, de la précision du résumé :

- premièrement, les **contenus libres d'accès**, qui ne servent pas par définition à établir de manière certaine l'existence d'un droit à rémunération, et sous réserve que des protocoles aient été mis en place pour s'assurer de leur nature et du contrôle d'une éventuelle réserve des droits, ce qui relève en partie de la politique de conformité (voir *supra*) ;
- deuxièmement, les **contenus protégés - dont il est essentiel de rappeler qu'aucune exception ou législation n'autorise quelque usage que ce soit sans autorisation** - doivent faire l'objet d'une attention particulière, et la preuve doit alors être apportée qu'ils ont été moissonnés dans le respect des prérogatives des ayants droit.

Lors de son déplacement à Bruxelles, la mission d'information a pu comprendre que la Commission européenne était spécifiquement attachée au sens premier du terme « résumé », qui exclut selon elle **toute notion d'exhaustivité ou de complétude**. Cette approche est justifiée notamment par la nécessité de ne pas handicaper une industrie européenne de l'IA encore

balbutiante, alors que plusieurs grands modèles existent déjà sur le marché (voir *infra*). À ce jour cependant, la Commission européenne n'a toujours pas fourni de modèle.

La mission d'information rappelle son attachement à l'élaboration d'un modèle respectueux des finalités du RIA, et donc en mesure de rassurer les ayants droit sur l'utilisation qui aura été faite - ou pas - de leurs données protégées.

À ce titre, elle partage pleinement les conclusions de l'avis politique adopté le 14 mai 2025 par la commission des affaires européennes du Sénat sur le code de bonnes pratiques en matière d'Intelligence artificielle à usage général¹ à l'initiative de Catherine Morin-Desailly et Karine Daniel, qui rappelle avec force qu'il appartient à la Commission européenne de « *respecter l'esprit et la lettre du règlement sur l'intelligence artificielle et d'affirmer avec force son attachement à la défense du droit d'auteur, des industries culturelles et de l'information, de la liberté des médias et des droits des journalistes* ».

C. EN AVAL : QUEL STATUT JURIDIQUE POUR LES ŒUVRES GÉNÉRÉES PAR UNE IA ?

Le statut juridique des œuvres générées par IA pose des questions pratiques auxquelles ni la législation, ni la jurisprudence, n'ont encore pleinement répondu. **L'IA interroge en effet la notion de droit d'auteur telle qu'elle a été définie depuis le XVIII^e siècle à l'initiative de la France.**

1. La définition classique du droit d'auteur

De manière classique, le droit d'auteur bénéficie à une **personne**, dont la création doit faire preuve d'**originalité**.

L'article L. 111-1 du code de la propriété intellectuelle attribue à l'auteur des droits sur l'œuvre produite : « *L'auteur d'une œuvre de l'esprit jouit sur cette œuvre, du seul fait de sa création, d'un droit de propriété incorporelle **exclusif** et opposable à tous* ». Il convient de noter que ce droit ne peut être attribué à une personne morale, comme l'a rappelé la chambre civile de la Cour de cassation dans l'arrêt n° 13-23.566 du 15 janvier 2015 : « *une personne morale ne peut avoir la qualité d'auteur* ».

La condition d'**originalité** ne figure pas dans le code de la propriété intellectuelle, mais à l'article 2.3 de la Convention de Berne du 9 septembre 1886 pour la protection des œuvres littéraires et artistiques. Cette notion a donné lieu à une vaste et ancienne jurisprudence, synthétisée par un arrêt n° 76-15.367 de Chambre civile de Cour de cassation

¹https://www.senat.fr/fileadmin/cru-1750816532/Commissions/Affaires_europeennes/Fichiers/Avis_politiques/AP_code_bonnes_pratiques_ADOPTE_def.pdf

du 6 mars 1979 « *si les œuvres de l'esprit sont protégées quels qu'en soient le genre, la forme d'expression, le mérite ou la destination, c'est à la condition qu'elles soient originales* ». Dans un arrêt du 16 juillet 2009, la Cour de Justice de l'Union européenne (CJUE) a défini l'originalité comme « *la création intellectuelle propre à son auteur* ». De fait, toute œuvre bénéficie d'une forme de **présomption d'originalité**, à partir du moment où elle reflète les choix de son auteur et porte sa vision du monde, indépendamment de tout jugement sur la qualité ou l'intérêt.

La reconnaissance de ce droit n'impose au demeurant **aucune formalité particulière**. L'article L. 111-2 du code précité précise ainsi que : « *L'œuvre est réputée créée, indépendamment de toute divulgation publique, du seul fait de la réalisation, même inachevée, de la conception de l'auteur* », alors que l'article L. 113-1 indique que : « *La qualité d'auteur appartient, sauf preuve contraire, à celui ou à ceux sous le nom de qui l'œuvre est divulguée.* »

2. Le défi juridique posé par l'IA

Les algorithmes utilisés par l'IA interrogent la notion de droit d'auteur d'une manière qui n'a pas été réellement anticipée.

Les créations générées par IA peuvent en effet être de deux natures : soit **assistées par IA**, soit **intégralement générées par celle-ci** avec une intervention humaine minimale.

a) Les contenus assistés par IA

Dans le premier cas, l'IA agit d'une certaine manière comme le **prolongement de la volonté de l'auteur**, de la même manière qu'un appareil photographique agit comme un auxiliaire technique pour produire un cliché dans les conditions exactes voulues par son auteur. L'IA est alors un simple outil au service de la création, qui ne ferait que rendre plus simple, plus efficace ou plus rapide un processus créatif. Le droit d'auteur peut alors sans difficulté être attribué à la personne à l'origine de la création.

Cependant, il existe une marge d'incertitude liée au fait **qu'il n'est actuellement pas possible de s'assurer que tel ou tel contenu a été intégralement généré par IA ou a simplement servi à la conception**.

Un exemple particulièrement frappant en a été donné avec l'interdiction faite par Amazon, pour un même auteur, de publier **plus de trois livres par jour** sous son propre nom sur la plateforme d'auto-édition *Kindle Publishing*. La plateforme, qui a constaté l'afflux massif de livres fortement soupçonnés d'avoir été produits à l'échelle industrielle par des IA, impose maintenant que les auteurs et éditeurs signalent si le livre a été généré par une IA, sans réelle possibilité de vérifier cette condition. De nombreux sites comme Bramewok proposent désormais d'assister les écrivains à rédiger leur livre,

cette aide pouvant aller jusqu'à la rédaction d'une simple commande comme « *Rédiger un roman policier qui ressemble à une œuvre X ou Y* ».

Le cas échéant, **les tribunaux peuvent aller assez loin dans l'exigence d'une intervention humaine**. Le Bureau américain des brevets a ainsi rendu un jugement le 5 septembre 2023, qui a refusé d'accorder le bénéfice de la propriété intellectuelle à l'artiste Jason Allen pour son œuvre graphique *Théâtre d'opéra spatial*, alors que l'auteur indique avoir rédigé plus de 624 instructions (prompts) sur l'IA « Midjourney ».

b) Les contenus intégralement générés par IA

En apparence, le cas d'une œuvre dont il est **dès l'origine** reconnu qu'elle a été générée par une IA avec une intervention humaine minimale est plus simple en termes de propriété intellectuelle, ne serait-ce que parce que **l'identité** du bénéficiaire des droits serait alors incertaine : le détenteur de la licence, les développeurs, la société propriétaire de l'IA ?

La législation comme les jurisprudences laissent à cet égard peu de doutes.

La nécessité que l'auteur soit bien une personne « humaine » a ainsi été confirmée tant par la justice européenne qu'américaine. Par exemple, l'inventeur américain Stephen Thaler, créateur d'une IA dénommée « DABUS¹ » qui vise **explicitement** à générer des œuvres avec une intervention humaine minimale, a déposé plusieurs demandes afin de faire reconnaître le statut d'auteur à son logiciel. Le tribunal du district de Columbia a finalement confirmé la décision de rejet du Bureau américain du copyright (US Copyright Office) le 18 août 2023, suivant un même jugement émis par la chambre des recours de l'Office européen des brevets le 27 janvier 2020. La chambre se base sur l'article 81 de la Convention sur le Brevet Européen du 5 octobre 1973 qui indique que « *La demande de brevet européen doit comprendre la désignation de l'inventeur* ». En conséquence, **la qualité d'inventeur comme d'auteur ne s'applique qu'à une personne dotée de la capacité juridique, ce qui n'est pas le cas d'un logiciel**.

Le Bureau américain du Copyright a consacré son rapport annuel de l'année 2025 à la question de l'IA. Les conclusions de ses travaux, remises le 29 janvier 2025², précisent que « *les œuvres produites par une IA générative peuvent être protégées par les lois du copyright seulement quand l'auteur humain a fixé les éléments expressifs suffisants. Cela peut inclure des cas dans lesquels une œuvre créée par un humain est perceptible dans la production de l'IA, ou dans lesquels un humain réalise des arrangements créatifs ou des modifications à cette production* ».

¹ « Device for the Autonomous Bootstrapping of Unified Information », en français : Dispositif pour l'auto-amorçage d'informations unifiées

² <https://www.copyright.gov/newsnet/2025/1060.html>

La question de l'**originalité** de l'œuvre produite par IA, si elle n'a pas encore donné lieu à des décisions de justice, pose également un problème presque philosophique. Par nature, l'IA agrège un grand nombre de données, et « imite » la production existante. Dans un propos tenu sur France Culture en 2023 rappelé par Le Monde¹, l'ancien président du Centre national de la Musique Jean-Philippe Thiellay pose la question : « *Est-ce qu'une intelligence artificielle aurait pu inventer « l'accord de Tristan » ? [en référence à l'opéra Tristan et Isolde (1865), de Richard Wagner, dont l'ouverture a constitué un jalon dans l'histoire musicale]. Est-ce qu'une IA aurait pu inventer les premiers morceaux de hip-hop ou le dernier morceau [de l'album *Rough and Rowdy Ways* de 2020] de Bob Dylan, qui dure dix-sept minutes ? Est-ce qu'on peut apprendre à la machine à être disruptive ?* »

Pour l'instant, **il semble qu'aucune œuvre explicitement générée par une IA n'ait franchi ce seuil, ce qui conforterait les humains dans l'idée que l'acte créatif « pur » est encore un apanage de l'humanité.**

Le précédent AlphaGo : l'IA créative ?

Les premières IA ont été entraînées sur des jeux. Il est très rapidement apparu qu'une IA jouant aux dames face à un bon joueur ne pouvait parvenir qu'à un match nul. Pour le jeu d'échecs, l'ordinateur Deep Blue a fini par remporter un tournoi en 6 manches contre le champion du monde Gary Kasparov en 1997, avec deux victoires.

Le cas du jeu de Go a pour sa part été longtemps considéré comme un défi ultime pour l'IA. Si les règles sont simples, il s'agit d'un jeu abstrait, qui se caractérise par un nombre de coups possibles extrêmement élevé et nécessite une grande intuition. En 2015, le consensus général était qu'il faudrait des décennies pour qu'un ordinateur puisse triompher des meilleurs spécialistes mondiaux.

¹ Le Monde, 15 mars 2025, « L'IA, menace réelle ou fantasmée pour les artistes ? » - https://www.lemonde.fr/idees/article/2025/03/15/l-ia-menace-reelle-ou-fantasmee-pour-les-artistes_6581226_3232.html

Un match est cependant organisé en 2016 à Séoul entre l'IA développé par Google « Alpha Go » et l'un des meilleurs joueurs de l'histoire, le coréen Lee Sedol. Il apparaît assez rapidement que la machine domine largement son adversaire. Lors de la dernière manche, elle réalise des coups jamais vus sur un Goban que même les programmeurs ne s'expliquent pas. Selon les mots du champion coréen, la machine aurait su faire preuve de *créativité*. Depuis cette date, les logiciels de Go qui utilisent une IA développée ne peuvent plus être battus par des humains.

Quelques pays ont établi une législation distincte pour accorder la protection du droit d'auteur à l'IA.

La Grande-Bretagne a élaboré dès 1988 une législation spécifique avec le Copyright, Designs and Patents Act 1988 (CDPA) UK¹, qui précise que le détenteur des droits « *doit être considéré comme la personne par laquelle les dispositions nécessaires à la création de l'œuvre sont prises*² », la seule condition étant que le contenu doit être original. Cependant, édictée à une époque où l'intelligence artificielle était loin de son développement actuel, cette disposition pourrait ne plus être d'actualité et n'a pas encore fait l'objet d'une décision des tribunaux britanniques. **L'Ukraine a fait évoluer sa législation en décembre 2022 pour accorder la protection du droit d'auteur à l'ensemble des créations issues de l'IA, qu'elles soient originales ou non.**

Il n'en reste pas moins que, face à la quantité, sinon la qualité, des œuvres produites par l'IA, **le cadre juridique paraît à ce jour peu adapté, et largement dépendant de décisions prises progressivement par les juridictions et les instances de régulation.**

La position constante des personnes entendues durant la mission d'information est que les contenus générés par IA ne devraient pas pouvoir bénéficier du régime de la propriété intellectuelle. Conscient des incertitudes qui entourent le statut des données sortantes ou « extrants », le CSLPA a confié le 20 juin 2025 une mission relative à la protection des contenus générés avec le recours à l'IA générative³ à la professeure Alexandra Bensamoun. **La commission suivra avec grand intérêt ses conclusions, attendues à l'été 2026.**

¹ <https://www.aoshearman.com/en/insights/ownership-of-ai-generated-content-in-the-uk>

² "shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken"

³ <https://www.culture.gouv.fr/fr/nous-connaître/organisation-du-ministère/Conseil-supérieur-de-la-propriété-littéraire-et-artistique-CSPLA/travaux-et-publications-du-cspla/missions-du-cspla/le-cspla-lance-une-mission-relative-a-la-protection-des-contenus-generes-avec-le-recours-a-l-ia-generative>

Le droit d'auteur s'est largement construit de manière jurisprudentielle, à travers les critères souvent souples des lois et traités internationaux. **Il est essentiel de réaffirmer le lien indéfectible qui doit lier le bénéfice de la protection intellectuelle à un ou des auteurs humains.**

*

* *

L'IA est allée bien plus rapidement que la législation, et se trouve aujourd'hui freinée dans son développement et ses perspectives par de fortes incertitudes juridiques, qui tiennent aussi bien aux éléments indispensables à sa conception qu'au statut des œuvres qu'elle génère.

IV. TRANSPARENCE ET RÉMUNÉRATION DANS L'INTÉRÊT DE TOUS

A. L'INSOUTENABLE GRATUITÉ DES DONNÉES

La très faible valeur, voire la totale gratuité, des données utilisées pour entraîner et spécialiser les modèles d'IA, semble aujourd'hui faire l'objet d'un large consensus de la part des sociétés de la tech.

Pour fonctionner, ces modèles nécessitent des puces de haut niveau, pour l'essentiel fournies par l'entreprise américaine Nvidia, des ingénieurs et des quantités considérables d'énergie, autant d'éléments dont nul ne revendique la gratuité. Dès lors, **pourquoi les données seraient-elles le seul élément indispensable de la chaîne de valeur pour lequel personne ne souhaite verser une rémunération ?**

La mission d'information a pu recueillir plusieurs justifications à cet état de fait qui constitue aujourd'hui **la nouvelle doxa des entreprises de la tech**, mais également un paradoxe. Les arguments avancés sont de trois ordres :

- **tout d'abord**, la nature même des données rendrait impossible ou trop complexe à établir toute forme de rémunération ;
- **ensuite**, les œuvres générées par IA seraient comparables aux créations de l'esprit humain, où les données ne seraient que des sources d'inspiration ;
- **enfin**, la rupture technologique introduite par l'IA serait si majeure qu'elle justifierait pour les États un aménagement des règles existantes.

1. Selon la tech, des données gratuites par nature

Si les données sont un élément indispensable à l'entraînement des IA, leur nature même les singularise sous plusieurs aspects des puces ou des compétences technologiques.

a) Les données culturelles sont facilement accessibles

Le premier argument revient sur la **facilité d'accès des données**. Le développement massif de l'économie numérique ces vingt dernières années a en effet conduit à la digitalisation de la quasi-intégralité des contenus culturels dans le monde, dorénavant disponibles sur Internet. Comme précisé dans la partie II du présent rapport, les fournisseurs ont pu accéder à de gigantesques jeux de données, obtenus directement à l'aide de robots d'indexation ou en utilisant gratuitement des bases déjà constituées comme celles de l'américain Common Crawl, qui rassemble en 2025 trois milliards de pages pour plus de

400 Terabit¹. Le développement des IA n'aurait pas été possible sans cette masse de données accessibles.

Dans les premiers temps de l'IA, cette facilité d'accès, y compris pour les contenus protégés, a donc instillé l'idée que les données n'avaient pas de valeur, puisqu'aucune démarche et aucun paiement n'était requis pour y accéder.

Une analogie : le marché de la musique enregistrée dans les années 2000

Ce mouvement n'est pas sans rappeler celui observé au tournant des années 2000 avec l'irruption des procédés de piratage à grande échelle par les réseaux pair-à-pair. Le marché de la musique enregistrée, qui a été le premier à en être victime, a ainsi vu son chiffre d'affaires en France divisé par trois entre 2002 et 2015, passant de 1,3 milliard d'euros à 426,4 millions.

L'industrie musicale a cependant su s'adapter à ce nouveau paradigme, grâce aux sites de *streaming* qui ont émergé comme Spotify ou Deezer et ont proposé des catalogues complets, ergonomiques, pour un abonnement d'un montant accessible au plus grand nombre. S'il est toujours possible de se procurer de la musique de manière illégale, cette pratique s'est cependant fortement contractée car elle nécessite des compétences techniques et une perte d'ergonomie que la majorité du public n'est pas prêt à fournir.

Il est regrettable de constater que les ayants droit n'ont pris conscience que trop tardivement du problème et n'ont pas su mettre en place des mécanismes de protection efficaces pour les œuvres. **Il est maintenant avéré que l'ensemble des contenus en ligne a d'ores et déjà été moissonné par les entreprises de la tech pour l'entraînement de leurs IA.** Cette utilisation a été justifiée *a posteriori* par une argumentation juridique autour des concepts d'exception TDM en Europe et de « *fair use* » aux États-Unis (voir *supra*), mais la légalité de cette interprétation est l'objet de débats et devra probablement être jugée devant les tribunaux des deux côtés de l'Atlantique.

Les données ont donc pu être utilisées car largement disponibles et accessibles, y compris à l'aide d'arguments juridiques contestables, une caractéristique que ne partagent pas d'autres éléments indispensables de la chaîne de production d'une IA comme les puces informatiques avancées, l'énergie, les ingénieurs développeurs, pour lesquels la gratuité n'a jamais été envisagée.

b) Les données culturelles ne vaudraient rien en raison de leur masse

Corollaire de leur accessibilité et de leur gratuité, les données ont pu être utilisées en très grande quantité. Il n'existe pas d'estimation fiable et incontestée de la masse requise pour entraîner une IA, mais les spécialistes s'accordent pour dire qu'une quantité énorme est nécessaire, surtout pour les

¹ Un terabit représente 10^{12} bits

modèles de fondation d'IA générative qui peuvent répondre à une grande variété de requêtes dans les domaines de l'écrit, de l'image ou de la vidéo.

La masse même de ces données conduirait *de facto* à **une valeur marginale quasi nulle** pour chacune considérée indépendamment. Prise isolément, la valeur d'un livre, d'un film, d'une photo, serait **infime** et aucun élément n'est en soi indispensable pour entraîner la machine, à l'exception de données très spécialisées dans certains secteurs.

Comment estimer la valeur des données ?

Dans son rapport réalisé au nom du CSLPA sur la rémunération des contenus culturels utilisés par les systèmes d'intelligence artificielle remis en mai 2025¹, Joëlle Farchy signale l'existence de deux procédures destinées à quantifier la valeur des données :

- le « *leave one out* », qui consiste à comparer les résultats de sortie d'un modèle entre une version entraînée avec certaines données, et le même modèle entraîné sans ces données ;

- la « *Shapley value*² », qui consiste, à l'inverse, à ajouter les données au modèle au fur et à mesure de manière incrémentale, et de tester l'amélioration de chaque itération sur le résultat final.

Ces deux méthodologies, très intuitives, ont cependant de sérieuses limites : d'une part, elles supposent d'entraîner plusieurs fois le même modèle dans des conditions différentes, ce qui a un coût d'autant plus prohibitif que de nombreuses données sont utilisées, d'autre part, elles ne sont pas totalement fiables compte tenu de la présence dans les bases de données de multiples contenus non étiquetés, et donc susceptibles d'être présents sans que l'expérimentateur ne le sache.

Une autre approche, plus simple à mettre en œuvre, consiste enfin à mesurer la similarité entre les contenus générés par l'IA et les données d'entraînement, et donc d'attribuer une valeur plus ou moins grande à telle ou telle donnée.

Enfin, certains chercheurs ont également testé un modèle en apposant un filigrane spécifique à chaque donnée entrante, ce qui permet de retracer les sources principales utilisées par l'IA pour une de ses productions.

c) *La rémunération serait trop complexe à établir faute d'interlocuteurs*

La valeur réputée infime de chaque donnée est souvent mise en avant par les fournisseurs d'IA pour justifier l'absence de rémunération. Cet argument est d'autant plus avancé que les différents secteurs culturels présentent une très grande hétérogénéité dans leurs modalités d'organisation

¹ <https://www.culture.gouv.fr/fr/nous-connaître/organisation-du-ministère/Conseil-supérieur-de-la-propriété-littéraire-et-artistique-CSPLA/travaux-et-publications-du-cspla/missions-du-cspla/mission-relative-a-la-rémunération-des-contenus-culturels-utilisés-par-les-systèmes-d-intelligence-artificielle>

² Du nom de Lloyd Shapley, qui a démontré en 1953 un important résultat en théorie des jeux sur la répartition des gains dans un jeu coopératif.

et de représentation. Ainsi, si **l'ensemble** des œuvres d'un genre (musique, édition...) représente une très grande valeur pour l'entraînement de l'IA, et pourrait éventuellement faire l'objet de négociations avantageuses pour les ayants droit, tel n'est pas le cas si les créateurs souhaitent entamer des négociations sectorielles sans atteindre une taille critique suffisamment significative.

Les fournisseurs d'IA renonceraient donc à rémunérer les contenus culturels **faute de l'existence d'un marché solide et fiable leur permettant d'acquérir rapidement les droits sur les œuvres en limitant les coûts de transaction.**

La réflexion sur la création d'un tel « marché de la donnée », dont le rapport précité de Joëlle Farchy étudie les conditions d'existence, serait donc un préalable à toute forme de rémunération, dans un secteur économique qui se développe extrêmement rapidement et soumis à une très forte concurrence.

2. La nature radicalement nouvelle des contenus culturels générés par l'IA

a) Des contenus totalement nouveaux

Une partie de la controverse sur l'usage des données par l'IA tient au fait que les créations générées par les modèles ne seraient que des répliques ou des imitations d'œuvres existantes, d'où l'expression « **quasi-œuvres** » déjà mentionnée. Dans ce cas, l'IA ne ferait que produire des contrefaçons assez élaborées pour rendre moins facilement discernables les sources.

Les concepteurs de modèle d'IA ne s'accordent pas sur cette vision. Selon eux, l'IA ne modifie pas des œuvres existantes, mais **s'en inspire**, en extrayant les schémas, les régularités, le sens général, avant de proposer, à l'aide de mécanismes d'inférence, un résultat original.

b) Préserver la qualité et la diversité des contenus synthétiques

Face à la volonté de réguler l'accès aux données, **les principaux fournisseurs soulignent les risques d'une réduction des bases de données qui aurait des répercussions sur la diversité et la qualité des productions générées.** Ainsi, si aucun contenu européen n'a été utilisé afin d'entraîner une IA, les contenus générés demeureront dépendants des sources utilisées et véhiculeront une vision du monde qui ne serait pas la nôtre. Cette crainte est particulièrement sensible dans le domaine de l'information. En effet, une IA qui aurait été entraînée sur des données fausses ou complotistes fournira des informations de même nature. Ainsi, en mai 2025, le moteur IA du réseau X Grok a spontanément fourni, en réponse à des questions pourtant éloignées, des réponses accréditant l'existence en cours d'un « génocide blanc » en Afrique du Sud, ce qui semble plus traduire les obsessions de son propriétaire Elon Musk qu'une réalité avérée.

Plus largement, les États qui feraient le choix de limiter l'accès à leurs données culturelles ou informationnelles agiraient **contre leurs propres intérêts** en effaçant des productions synthétiques toute trace de leur culture.

Il y aurait donc **un intérêt civilisationnel à accorder un libre accès aux données pour un pays qui souhaiterait continuer à faire rayonner sa culture.**

3. Une technologie de rupture qui justifie une évolution des règles

a) Obtenir un avantage concurrentiel

Perçu comme une des grandes technologies qui façonneront le futur, **le développement de l'IA est devenu un enjeu majeur pour assurer la souveraineté économique et politique des États.** À ce titre, la création d'un environnement favorable à son développement est une priorité, plusieurs fois réaffirmée aussi bien par les États-Unis que par l'Europe.

La Commission européenne cherche ainsi à développer une stratégie européenne de l'IA, formalisée dans plusieurs documents¹, le dernier datant du mois d'avril 2025 pour « *promouvoir l'IA à l'échelle du continent*² ». De manière significative, la question des données et du respect de la propriété intellectuelle n'y est mentionnée qu'une seule fois, alors que l'attention porte prioritairement sur le développement des infrastructures techniques nécessaires. *De facto*, et comme le souligne dès son préambule la section « Copyright » de la troisième version du code de bonne pratique prévu à l'article 53 du RIA (voir *supra*), « *la conformité des engagements doit être proportionnée et proportionnelle à la taille et aux capacités des fournisseurs, en prenant en compte les intérêts des petites entreprises* ». **La stratégie européenne se comprend donc comme visant à créer l'environnement le plus favorable possible au développement de l'IA pour faire de l'Europe le « continent leader » en la matière.**

À l'occasion de son déplacement à Bruxelles, la mission d'information a eu l'occasion de mieux comprendre **la position de la Commission européenne. Une élévation trop significative des standards en matière de protection des droits d'auteur ferait, selon elle, peser un risque compétitif sur le développement des IA européennes,** alors même que les États-Unis et la Chine encourageraient de manière agressive leurs propres technologies.

Cette position très proche de celle défendue par la France durant les négociations sur le projet de RIA, met donc en opposition les défenseurs des droits d'auteur et les entreprises européennes.

¹ <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>

² <https://digital-strategy.ec.europa.eu/en/library/ai-continent-action-plan>

b) Une stratégie assumée de puissance

Les enjeux liés à l'IA dépassent cependant dorénavant le secteur économique. Les États ont ainsi élevé le développement de ces technologies au rang de **priorité nationale** et mettent en avant le risque de déclassement pour justifier des évolutions de la législation.

Dans un mémorandum adressé le 13 mars 2025 à l'administration américaine, **la société OpenAI se livre ainsi à un plaidoyer pour le développement des IA américaines**, mêlant des arguments économiques et géopolitiques. La liberté qui doit être laissée aux entreprises américaines n'est pas seulement présentée comme un impératif économique, mais également comme **une priorité au titre de l'intérêt national**, la République Populaire de Chine étant nommément visée comme le principal ennemi à combattre.

De manière significative, **ce document ne mentionne à aucun moment l'Europe comme un concurrent potentiel**, mais alerte sur les risques que ferait peser sur le développement des IA américaines la mise en œuvre d'une régulation sur les données. OpenAI encourage ainsi le gouvernement américain à « *façonner les discussions politiques internationales autour des droits d'auteur, et à travailler à empêcher les pays moins innovants d'imposer leur régime juridique aux entreprises américaines d'IA et à ralentir leurs progressions* ». Cette allusion évidente au RIA, qui exprime au passage un jugement que l'on pourrait juger méprisant à l'égard des pays européens, souligne que **la question de l'accès aux données dépasse le simple cadre de la concurrence entre entreprises**.

Il existe donc de nombreux arguments qui plaideraient en faveur de la quasi-gratuité des données, allant de leur nature à des considérations géopolitiques. **Avancés à tour de rôle devant la mission d'information, ils illustrent les difficultés auxquelles sont confrontés les ayants droit, accusés tour à tour de ne pas comprendre les spécificités de l'IA, de surévaluer la valeur de leurs œuvres dans ce modèle économique, et d'entraver le développement d'une technologie non seulement prometteuse, mais également incontournable pour le futur.**

B. UNE SITUATION PRÉJUDICIALE À TOUS

L'absence de rémunération et de transparence sur les données présente en réalité plusieurs risques très significatifs, non seulement pour les créateurs, mais également pour les fournisseurs d'IA, singulièrement en Europe.

1. Des menaces pour les créateurs

a) Un modèle économique fragilisé

Comme cela a été mentionné précédemment, les créateurs et toute la chaîne de valeur qui gravite autour d'eux (producteurs, éditeurs etc...) sont les premières victimes de l'usage sans rémunération des œuvres. Ils subissent ainsi un double effet : **en amont**, leurs productions sont utilisées sans rémunération, **en aval**, l'utilisation qui est faite de leurs œuvres nourrit un système qui à terme proposera des œuvres **en concurrence directe avec leurs propres productions**.

Plus grave, **l'absence de rémunération constitue un signal extrêmement négatif adressé par les fournisseurs au monde de la création, accréditant l'idée que les œuvres n'auraient en elles-mêmes aucune valeur**, au regard des éléments rappelés précédemment. Ce raisonnement ne tient cependant pas la route compte tenu de la dépendance des modèles d'IA à ces mêmes données, sans lesquelles ils ne peuvent produire aucun résultat.

La rémunération constitue donc *a minima* une reconnaissance de la valeur des données utilisées.

b) Donner aux créateurs les moyens de s'adapter à l'IA

Les relations entre IA et création ne se limitent pas à une forme de prédation et de concurrence déloyale. Les nombreux outils proposés par l'IA sont également des leviers qui offrent de nouvelles palettes d'expression artistique, comme cela a été expliqué dans la partie II du présent rapport.

Or l'intégration de nouvelles technologies suppose **des actions d'acquisition de matériels et de formation aux nouveaux usages**. Les effets spéciaux dans le cinéma ont pu se développer parce que les studios ont tout d'abord généré des revenus avec des productions classiques, revenus qui ont permis un bond technologique pour les œuvres suivantes. Tel n'est pas le cas aujourd'hui, où **les revenus demeurent pour l'instant concentrés entre les mains des seuls fournisseurs d'IA**.

Il est donc essentiel que les revenus générés par l'IA soient aussi utilisés pour permettre aux créateurs de s'adapter à cette nouvelle révolution technologique dont ils doivent aussi être acteurs.

2. Une double incertitude pour les entreprises technologiques

Si les avantages de la gratuité des données semblent évidents pour les fournisseurs, il présente cependant à terme des risques significatifs pour leur développement, singulièrement pour les entreprises européennes.

a) *Les risques communs à tous les fournisseurs d'IA*

(1) Des incertitudes juridiques lourdes

Les entreprises de technologie soutiennent que l'utilisation des données est couverte, en Europe, par l'exception TDM – sans pour autant en respecter les termes, voir *supra* – et aux États-Unis, par le « *fair use* ».

Dans sa lettre précitée du 13 mars 2025, OpenAI en fait un point central : « *Les États-Unis ont tellement de startups dans le domaine de l'IA, attirent tellement d'investissements, et ont réalisé tellement d'avancées largement parce que la doctrine du fair use promeut le développement de l'IA. Dans d'autres marchés, les règles rigides de la propriété intellectuelle freinent l'innovation et l'investissement¹* ».

Cependant, il n'existe pas encore de consécration devant les tribunaux américains de cette interprétation. Bien au contraire, plus de 60 contentieux sont actuellement engagés dans plusieurs secteurs pour contester l'application de la doctrine du *fair use* pour l'entraînement des IA. Ce débat ne concerne pas uniquement le secteur créatif. La plateforme Reddit a ainsi porté plainte contre la société Anthropic, créateur de l'IA « Claude », pour avoir utilisé sans autorisation les propos tenus publiquement sur son site. Reddit avait d'ailleurs conclu des accords avec plusieurs fournisseurs d'IA pour l'usage de ses contenus, dans des conditions permettant de protéger la vie privée de ses utilisateurs.

La situation est sensiblement la même en Europe. Pour la France, le Syndicat national de l'édition (SNE), la Société des gens de lettres (SGDL) et le Syndicat national des auteurs et des compositeurs (SNAC) ont ainsi saisi le tribunal judiciaire de Paris contre la société Meta, propriétaire de l'IA Llama, accusée d'utiliser leurs œuvres sans leur consentement. D'autres actions sont également en cours.

Les fournisseurs d'IA sont donc engagés dans des contentieux juridiques des deux côtés de l'Atlantique, qui mettront probablement plusieurs années à être résolus par un recours aux juridictions supérieures. **Or l'impact pour elles n'est pas négligeable. En effet, il n'existe pas de possibilité technique de « désentraîner » une IA.** Si un jugement devait contraindre une entreprise à « retirer » un contenu, **elle devrait *de facto* bâtir en totalité un nouveau modèle, pour un coût extrêmement élevé.** Le risque juridique est d'autant plus important que le droit américain assortit les jugements de dommages et intérêts qui peuvent être très élevés.

Face à une incertitude juridique potentiellement mortelle, les fournisseurs d'IA auraient donc un intérêt majeur à régler en amont leurs relations avec les ayants droit, ce qui permettrait de « solder le passé » et d'ouvrir des perspectives de coopération mutuellement profitables.

¹ *America has so many AI startups, attracts so much investment, and has made so many research breakthroughs largely because the fair use doctrine promotes AI development. In other markets, rigid copyright rules are repressing innovation and investment.*

(2) Vers l'effondrement des modèles ?

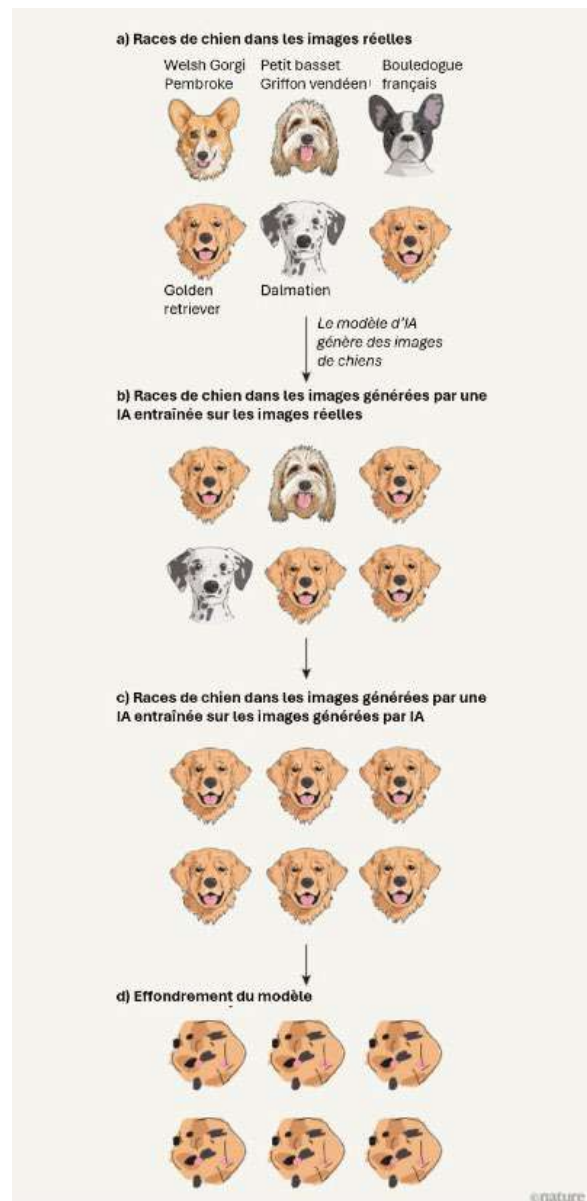
Les modèles d'IA ont utilisé d'énormes quantités de données pour parvenir à leur niveau actuel. De plus en plus de contenus présents en ligne sont désormais générés par des IA. Ainsi, environ 20 % des morceaux proposés par les services de *streaming* seraient totalement « synthétiques ».

En première analyse, une forme de « cercle vertueux » pour l'IA pourrait donc être engagée : des contenus synthétiques viendraient alimenter les bases d'entraînement pour perfectionner les IA. Cette thèse est cependant battue en brèche. Comme rappelé précédemment, plusieurs études récentes¹ ont ainsi montré que les modèles d'IA ont tendance à subir une forme de **dégénérescence** quand la quantité de données synthétiques utilisée pour l'entraînement dépasse un certain seuil. Les nouveaux contenus générés sont d'une qualité de plus en plus faible, avec des erreurs qui se reproduisent et s'accroissent au fil des itérations. *In fine*, le modèle est menacé d'effondrement : il ne produit plus que des résultats sans pertinence.

Dans son rapport précité, Joëlle Farchy propose une illustration particulièrement parlante de l'appauvrissement et finalement, de l'effondrement d'une IA entraînée sur des contenus synthétiques.

¹ Par exemple : Ilya Shumailov et alii, *Nature*, juillet 2024 : AI models collapse when trained on recursively generated data - <https://www.nature.com/articles/s41586-024-07566-y>

Illustration de l'effondrement d'une IA



Les modèles d'IA ont donc besoin d'être sans cesse alimentés de contenus humains « originaux ». Au-delà des considérations mathématiques et statistiques qui conduisent à prédire l'effondrement, **les requêtes en lien avec l'actualité sont particulièrement sensibles**. Une IA doit pouvoir disposer de sources fiables pour répondre à une interrogation sur le résultat d'une élection, un résultat sportif ou un n'importe quel événement récent, et ne peut pas se fier aveuglément aux publications des réseaux sociaux.

Il est donc dans l'intérêt des fournisseurs d'IA de maintenir des secteurs de la création et de l'information puissants et de qualité pour ne pas appauvrir et détruire ses propres modèles. Or la prolifération actuelle de contenus synthétiques sur Internet, pour la plupart sans indication de leur origine, pose la question de la **viabilité à terme des IA si les bases constituées pour leur entraînement ne sont pas précisément sourcées**.

3. L'Europe, « idiot utile » des Américains ?

L'Europe affiche de grandes ambitions en matière de développement des IA et déploie une politique active de promotion de ces technologies dans l'espoir de faire de l'Europe « le continent de l'IA ». Elle semble cependant emprunter le chemin d'un **alignement sur les pratiques américaines**, jugées plus efficaces, dans l'espoir que des entreprises européennes puissent en tirer profit.

Le précédent de la LCEN

La volonté d'aligner le droit européen sur le droit américain s'est déjà manifestée avec les législations adoptées pour faire face à l'essor des technologies numériques au début des années 2000.

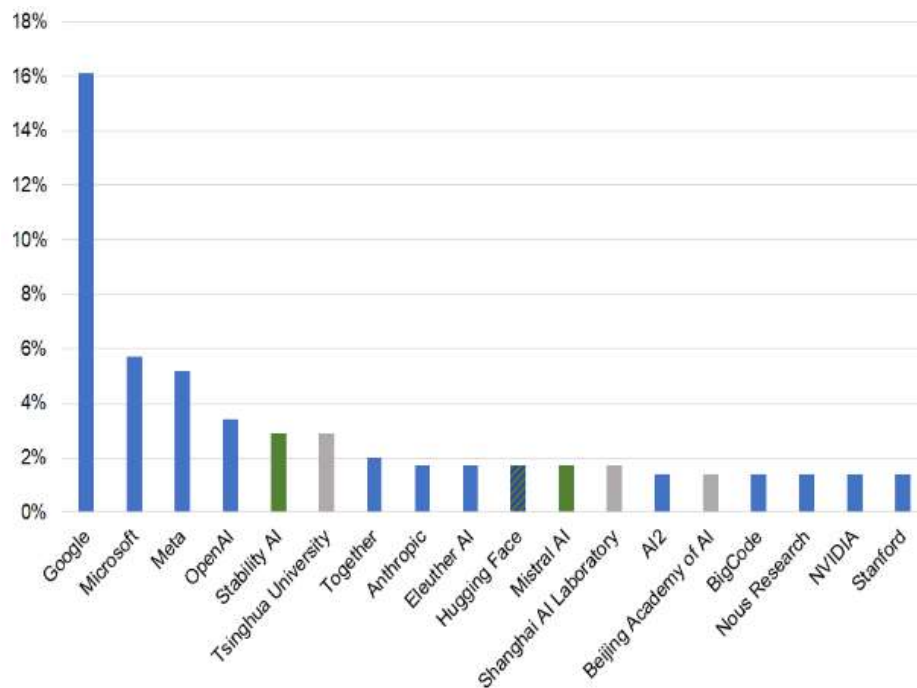
La directive 2000/31/CE du 8 juin 2000 a ainsi cherché à favoriser le développement des grandes entreprises européennes dans le secteur numérique, en reprenant pour l'essentiel le Digital Millennium Copyright Act adopté aux États-Unis en 1998. Transposé en droit français par la loi du 21 juin 2004 pour la confiance dans l'économie numérique (LCEN), la directive a en particulier introduit le statut d'hébergeur, qui figure à l'article 6-I-2 de la LCEN. Ce statut exonère les plateformes de toute responsabilité pour les contenus qui y sont publiés par les usagers, sauf cas de signalement. À l'usage, il s'est révélé extrêmement favorable pour les entreprises américaines, sans permettre à l'Europe de développer ses propres technologies¹.

Notre continent ne dispose cependant pour l'instant pas d'avantages en matière de prix de l'énergie, de marché des capitaux et de compétences technologiques dans le secteur des puces. De fait, le marché est actuellement très largement dominé par les États-Unis, comme le souligne l'étude de la Direction générale du Trésor de décembre 2024².

¹ Catherine Morin-Desailly a déposé le 27 septembre 2018 une proposition de résolution adoptée par le Sénat pour remettre en cause ces statuts d'hébergeur : <https://www.senat.fr/dossier-legislatif/ppr17-739.html>

² La chaîne de valeur de l'intelligence artificielle : enjeux économiques et place de la France <https://www.tresor.economie.gouv.fr/Articles/54e356a7-2319-43d3-bb84-361430cc82e6/files/a66dc035-21da-4f10-9997-9516ffa7cc3f>

Part des différentes organisations dans le nombre total de modèles de fondation entre le 1^{er} janvier et le 4 octobre 2024



Source : Étude de la direction générale du Trésor

À l’opposé, **l’Europe dispose de données-œuvres de très grande qualité**, exprimées en différentes langues et qui reflètent une large variété de sensibilités culturelles et artistiques. **Cela constitue, avec la formation des ingénieurs, notre plus grand avantage comparatif sur la Chine et les États-Unis.**

Dès lors, **un alignement « par le bas » sur la protection des droits d’auteur pourrait ne pas être un encouragement au développement des entreprises européennes, mais plutôt une manière pour des sociétés déjà en position dominante de conforter leur position pour un coût quasi nul.** L’administration américaine s’attaque d’ailleurs aux régulations européennes dans le domaine du numérique, qui vont de la protection des données personnelles – avec le RGPD adopté en mai 2018, depuis devenu un standard mondial – jusqu’aux Règlements européens de 2022 sur les services et les marchés numériques (DSA et DMA).

Protéger les données européennes en faisant respecter le droit d’auteur et la rémunération afférente est donc de nature à **faire émerger un modèle européen respectueux de la création, éthique, et alimenté par des données de qualité** dont l’usage ne souffrirait d’aucune incertitude juridique. **Cette pratique permettrait à l’Europe, dans la concurrence qui va s’engager autour de l’IA, de ne pas être « l’idiot utile » des Américains et d’éviter de se trouver prisonnière d’un féodalisme numérique qu’elle subit déjà.**

V. LES RECOMMANDATIONS DE LA MISSION D'INFORMATION POUR UN PARTAGE DE LA VALEUR ÉQUILIBRÉE ENTRE ACTEURS DE L'IA ET AYANTS DROIT CULTURELS, SOCLE SUR LEQUEL BÂTIR UNE TROISIÈME VOIE DE L'IA

Innovation de rupture, surtout depuis qu'elle est en capacité de générer des contenus ressemblant aux créations humaines, l'IA soulève des défis juridiques, économiques, culturels et éthiques à nos démocraties, même si on ignore encore largement ses limites.

À l'occasion de ses travaux, **la mission d'information a pu prendre la mesure aussi bien des inquiétudes légitimes des ayants droit culturels face à une IA sans garde-fous, que des potentialités offertes par cette nouvelle donne technologique pour le secteur de la création artistique.**

Elle est aussi pleinement consciente de l'enjeu stratégique de souveraineté que pose l'IA pour l'Union européenne, dans un contexte géopolitique marqué par l'émergence de « prédateurs », pour reprendre l'expression de Giuliano da Empoli¹, et par les attaques en règle de la nouvelle administration américaine contre toute forme de régulation des géants de la tech.

Convaincue de la nécessité de maîtriser les risques liés à l'IA, tout en favorisant son potentiel d'innovation, **la mission d'information souhaite poser les bases d'une régulation qui, au niveau européen, concilierait les intérêts de cette technologie et les préoccupations du secteur de la création.**

Pour ce faire, **la mission d'information formule huit grands principes** qui devraient, selon elle, **guider les parties prenantes sur la voie d'une rémunération appropriée pour l'utilisation des contenus culturels** par les acteurs de l'IA, permettant de **dépasser l'opposition stérile entre innovation et régulation** et d'**enclencher un cercle vertueux entre IA et création.**

En termes de méthode, la mission recommande une réponse graduée en trois étapes, qui laisserait d'abord sa place à la concertation entre les parties prenantes, puis qui définirait, en cas d'échec de celle-ci, les moyens juridiques nécessaires au respect des intérêts légitimes des ayants droit culturels.

A. L'ABSENCE PRÉOCCUPANTE D'UN MARCHÉ DES DONNÉES

À juste titre, les ayants droit culturels demandent à pouvoir profiter des dispositions du RIA pour **créer un marché réellement équilibré** où les contenus culturels seraient négociés et justement rémunérés.

¹ « L'heure des prédateurs » Giuliano da Empoli, 2025

Force est cependant aujourd'hui de constater qu'un tel marché ne peut pas se développer tant est grande **l'asymétrie d'informations entre les parties prenantes**. D'un côté, les fournisseurs d'IA moissonnent très largement les contenus culturels accessibles en ligne sans aucune considération pour le droit moral et patrimonial de leurs détenteurs légitimes, de l'autre, les ayants droit culturels ne peuvent que constater l'analogie entre les productions générées par l'IA et leurs œuvres, sans être en capacité de prouver l'usage qui en est fait.

L'économie de la fabrication de l'IA générative ne présente donc pas les caractéristiques d'un marché concurrentiel, **ce qui justifie pleinement, au regard de la théorie économique, une intervention publique**.

L'exemple pas tout à fait transposable des droits voisins des éditeurs et des agences de presse

Par le biais de la proposition de loi de David Assouline, le Sénat a été à l'origine de la première transposition en Europe de l'article 15 de la directive du 17 avril 2019, avec la promulgation le 24 juillet 2019 de la loi relative aux droits voisins des agences de presse et des éditeurs de presse. Ces droits permettent aux bénéficiaires d'obtenir une rémunération pour l'utilisation par les moteurs de recherche en ligne et les réseaux sociaux de citations et d'extraits issus de leurs publications. La détermination des montants dus se heurte cependant à des difficultés très similaires à celles observées aujourd'hui pour les œuvres utilisées par l'IA, à savoir l'asymétrie d'informations entre les éditeurs et agences de presse et les fournisseurs de service en ligne. Après des années de négociations très tendues¹ et infructueuses, suivis de contentieux où les éditeurs et les agences de presse ont dû acquitter des frais d'avocat considérables face à des géants du numérique à la surface financière incomparable, la solution est finalement venue de la régulation. Ainsi, l'Autorité de la concurrence a imposé à Google le 12 juillet 2021 et le 15 mars 2024, pour un montant total de 750 millions d'euros, pour non-respect de ses engagements, notamment le refus de communiquer des informations permettant d'établir le montant de la rémunération.

Si la décision du 15 mars 2024 est pour l'essentiel motivée par les conditions de négociation des droits voisins, l'Autorité a complété son jugement en soulignant que la société avait « *manqué à l'obligation de transparence en ne tenant pas informés les éditeurs et agences de presse de l'utilisation de leurs contenus par Bard [l'IA de Google, désormais dénommé Gemini]* ». L'Autorité souligne que Google n'a pas « *proposé de solution technique permettant aux éditeurs et agences de presse de s'opposer à l'utilisation de leurs contenus par Bard* », ce qui revient à donner une force effective au principe de l'*opt-out*.

¹ Une synthèse a été effectuée par le rapporteur pour avis des crédits de la presse dans son rapport sur le projet de loi de finances pour 2022 : <https://www.senat.fr/rap/a21-168-42/a21-168-426.html#toc77> et

Il n'est cependant pas certain que l'Autorité de la concurrence puisse exercer pour l'heure une telle action dans le domaine de l'IA. En effet, ce secteur n'est pas caractérisé, comme celui des moteurs de recherche, par une position dominante qu'il lui reviendrait alors de réguler. Dès lors, les espoirs qui ont pu être fondés sur une approche par le droit de la concurrence sont encore très incertains.

Très consciente des risques économiques et sociétaux que fait peser la révolution de l'IA, **la mission d'information souhaite promouvoir l'élaboration d'un cadre qui permettra aussi bien aux acteurs culturels de bénéficier largement et légitimement des retombées économiques qui lui sont dues, qu'aux entreprises européennes de l'IA de se hisser au tout premier rang mondial.**

Elle constate cependant que **la détermination du niveau de rémunération des contenus culturels utilisés par l'IA est une équation complexe à résoudre, pour laquelle aucune solution simple et partagée n'émerge à ce jour.** En effet, les données culturelles sont utilisées à différents niveaux dans la chaîne de valeur de l'IA, avec un degré de substituabilité variable, particularité qui n'existe dans aucun autre marché. Les analyses juridiques comme économiques disponibles ne peuvent s'appuyer que partiellement sur l'expérience, tant « l'objet IA » est par nature disruptif.

Dans ces conditions, **la mission d'information appelle à préserver du mieux possible un droit de la propriété intellectuelle solidement établi depuis plus de deux siècles,** et dont la France est à l'origine, **tout en travaillant son articulation avec la réalité d'un nouveau modèle économique** qui épouse mal les modèles connus et testés. Les modalités du futur modèle de rémunération qu'elle appelle de ses vœux doivent encore être approfondies par des travaux juridiques et économiques complémentaires à ceux déjà disponibles.

L'exemple allemand

De nombreux pays tentent de mettre en place des solutions juridiques solides. La diversité des approches témoigne à elle-seule de la complexité de la question. Le rapport précité pour le CSLPA de la professeure Alexandra Bensamoun mentionne ainsi une initiative allemande. La société pour les droits sur la représentation musicale et la reproduction mécanique (GEMA) a mis en place un modèle à deux composantes, qui prend en compte, d'une part, l'utilisation des données lors de la phase d'entraînement, d'autre part, la valeur générée par les productions de l'IA. L'avantage principal est d'associer étroitement le secteur musical aux éventuels succès des productions générées par les IA.

B. HUIT GRANDS PRINCIPES À RESPECTER DANS LE CADRE DE L'ÉLABORATION CONCERTÉE D'UN MODÈLE DE RÉMUNÉRATION DES CONTENUS CULTURELS UTILISÉS PAR L'IA

Face aux difficultés techniques du dossier, **la mission d'information tient à affirmer politiquement huit grands principes que toute mise en place d'un modèle de rémunération se devra de respecter.** Ces principes pourront utilement éclairer les travaux de la concertation lancée le 23 avril 2025, à l'initiative conjointe de la ministre de la culture et de la ministre déléguée chargée de l'IA et du numérique.

La stratégie d'action ministérielle pour une IA culturelle

Le 3 juillet 2025, le ministère de la culture a dévoilé sa stratégie d'action « pour une IA culturelle, responsable et souveraine ».

La mission d'information constate avec satisfaction que trois des cinq grands axes¹ de cette stratégie correspondent aux enjeux qu'elle a identifiés, à savoir :

- développer des IA et des usages responsables, en promouvant des modèles transparents, éthiques et frugaux, entraînés sur des données représentatives de la diversité culturelle et linguistique française ;

- garantir un modèle économique équitable, en protégeant les droits des créateurs, anticipant les mutations des métiers culturels, et assurant une juste rémunération face à l'essor des contenus générés par IA ;

- stimuler l'innovation dans la création artistique et l'offre culturelle, en soutenant les expérimentations artistiques, la recherche et les nouveaux services fondés sur l'intelligence artificielle.

Elle sera très attentive au déploiement de cette stratégie, notamment passées les échéances de la publication du code européen de bonnes pratiques et des résultats de la concertation nationale.

1. Premier principe : le droit à rémunération des ayants droit est légitime et incontestable

Le principe d'une rémunération pour l'ensemble des contenus culturels utilisés par les fournisseurs et déployeurs d'IA, quel que soit le moment du processus où ils sont utilisés, **est aussi légitime qu'incontestable.** Toutes les auditions d'économistes et de juristes menées par la mission d'information sont convergentes sur ce point.

¹ Les deux autres axes portent moins directement sur la création artistique.

Éléments indispensables de la chaîne de production d'une IA générative, les données servant à l'entraînement des modèles doivent faire l'objet d'une rémunération appropriée.

Les revenus ainsi générés doivent être mis à profit par le secteur culturel, en particulier pour **assurer la formation et l'accompagnement des professionnels, afin que ceux-ci ne subissent pas cette innovation de rupture mais en prennent le contrôle**, de la même manière que les évolutions technologiques dans la musique ont été financées par les ventes de productions plus anciennes.

Ce cercle vertueux, qui permet de nourrir le progrès, est pour l'instant rendu impossible par l'absence injustifiable de rémunération.

Recommandation n° 1 : Réaffirmer et garantir le droit à rémunération des ayants droit culturels pour l'utilisation de leurs contenus par les fournisseurs d'IA.

2. Deuxième principe : la transparence sur les données utilisées doit être garantie

Maintes fois évoquée, **la transparence sur les données utilisées par les fournisseurs d'IA apparaît comme la condition nécessaire – mais non suffisante – de la rémunération.** Elle est également un gage indispensable de confiance entre le secteur culturel et le secteur de la tech, tout comme une garantie sur la qualité des productions générées par l'IA.

Si on peut s'interroger sur le degré de granularité nécessaire et sur le caractère public des données, **la mission d'information estime que toutes les informations doivent être apportées sur les différentes étapes (entraînement, spécialisation, ancrage) au cours desquelles elles ont pu être utilisées.**

À ce titre, le « *résumé suffisamment détaillé* » mentionné à l'article 53 du RIA, s'il constitue une étape essentielle, ne sera pas en mesure d'apporter les éléments suffisants à l'établissement d'une rémunération représentative de l'usage. Il faut donc le considérer comme la partie « publique » d'une masse d'information qui devra nécessairement faire l'objet de négociations entre les parties prenantes, à la fois sur leur volume, mais également sur leur caractère éventuellement protégé.

Pour reprendre une analogie largement entendue durant les auditions, notamment utilisée par les professeures Alexandra Bensamoun et Joëlle Farchy dans leurs rapports respectifs pour le CSLPA, **si les ingrédients utilisés doivent être rendus publics, les quantités et le traitement peuvent, pour leur part, relever des négociations entre les parties prenantes.**

Recommandation n° 2 : Garantir la transparence complète des données utilisées par les fournisseurs d'IA.

3. Troisième principe : la rémunération doit être fonction des flux de revenus générés par l'IA

Il est encore difficile aujourd'hui d'estimer le niveau du « marché de l'IA » à moyen terme. Il existe différents secteurs dans lesquels cette technologie, dont on ignore les futures potentialités tout comme les futures limites, est déjà appliquée avec succès, comme la médecine ou l'analyse juridique.

Dans le domaine culturel, les hypothèses sont encore incertaines et reposent largement sur l'attrait que présenteront pour le public des productions synthétiques, moins onéreuses et disponibles en plus grand nombre, face à des œuvres purement humaines. Quelle sera la propension des consommateurs à payer pour détenir un livre, écouter une musique, visionner un film, s'ils ont été générés par des IA ? Pour l'instant, aucune œuvre culturelle entièrement synthétique ne semble avoir franchi la barrière du succès.

En tout état de cause, les IA génératives sont potentiellement en capacité de produire une infinité de contenus et ce, pendant encore de longues années. Elles auront donc **un usage en continu des données** qui servent à les nourrir.

Dans ce contexte, **la mission d'information estime qu'un modèle de rémunération équitable doit tenir compte de cette spécificité en associant, directement et en flux continu, les créateurs aux succès économiques de l'IA.**

Dès lors, la rémunération ne doit pas être marginale ou symbolique, mais **corrélée au chiffre d'affaires du secteur de la tech - fournisseurs comme déployeurs d'IA -**, afin de constituer **un levier de financement pour le secteur de la création**, qui plus est nécessaire à la poursuite de l'activité des IA qui ont un besoin continu de nouvelles données humaines de qualité.

De cette façon, **un cercle vertueux** pourra être mis en place : le succès de l'IA bénéficiera directement au secteur culturel dans son ensemble.

La mission estime donc nécessaire que la rémunération ne se résume pas à un simple paiement, pour solde de tout compte, d'un jeu initial de données, mais qu'elle soit conçue pour irriguer dans le temps les industries culturelles et créatives humaines.

Comme indiqué précédemment, les modalités de cette rémunération doivent être travaillées et discutées, la mission d'information souhaitant qu'un consensus entre les acteurs émerge à ce sujet.

Recommandation n° 3 : Définir des modalités de rémunération qui soient fonction des flux de revenus générés par les fournisseurs et déployeurs d'IA.

4. Quatrième principe : la création de bases de données harmonisées, aux conditions d'utilisation clairement définies, est un préalable indispensable à l'existence d'un marché de la donnée

Le préalable nécessaire à l'existence d'un marché attractif de la donnée, qui reconnaîtrait enfin aux ayants droit un pouvoir de marché qui n'aurait pas dû leur échapper, est la création de bases de données disponibles, larges, à défaut d'être complètes, qui comporteraient aussi bien les références que les fichiers, dans un format utilisable par les fournisseurs d'IA.

Par analogie, le secteur musical a pu se reconstruire après le choc provoqué par le piratage au début des années 2000 non pas en se fractionnant, mais en travaillant au nouveau modèle des plateformes de *streaming*. Sans écarter définitivement le piratage, ces dernières l'ont rendu moins attractif : pourquoi prendre le risque de frauder alors que, dans des conditions claires, une offre payante et ergonomique est disponible ?

La mission d'information estime donc que l'ensemble des filières créatives et la presse doivent d'urgence se mobiliser pour établir les conditions juridiques et techniques de la création de bases de données, en lien avec les acteurs de l'IA.

Il appartient donc au secteur culturel et à celui de la presse, dans leur diversité, de s'accorder sur des standards et des procédures communes, qui permettront *in fine* de garantir le droit d'auteur.

Recommandation n° 4 : Inciter le secteur culturel et celui de la presse à constituer des bases de données larges et de qualité, facilement exploitables par les fournisseurs, assorties de conditions d'utilisation précisément définies.

5. Cinquième principe : le passé doit être soldé

De son côté, le secteur de l'IA gagnerait à apurer les contentieux nés des conditions juridiques discutables de sa création. **Il fait peu de doutes aux yeux de la mission que l'exception TDM a été largement détournée de sa vocation initiale et qu'une vaste quantité de données protégées ont été moissonnées hors de tout cadre légal.**

Un même raisonnement pourrait trouver à s'appliquer aux États-Unis avec le « *fair use* ». Les contentieux en cours, qui devraient logiquement se développer dans les mois et les années à venir, devraient se solder par des

indemnités très élevées, que viendront encore faire grossir les frais juridiques de part et d'autre. Dans le pire des cas pour l'industrie de l'IA, des modèles entiers pourraient devoir être reconstruits faute de pouvoir en extraire telle ou telle donnée dont l'usage serait définitivement prohibé.

Dès lors, et comme première étape vers des relations apaisées, **la mission d'information recommande d'engager rapidement un dialogue pour parvenir à des accords mutuellement avantageux** qui éteindront les contentieux en cours, par exemple, par le biais de contrats rétroactifs.

Recommandation n° 5 : Parvenir à un règlement financier pour les usages passés des contenus culturels, afin de compenser les ayants droit culturels et sécuriser juridiquement les fournisseurs d'IA.

6. Sixième principe : un avantage comparatif doit être donné aux fournisseurs d'IA respectueux du cadre légal

Des bases de données larges et aux conditions d'accès clairement arrêtées seraient assurément d'une grande aide pour accélérer l'intégration du monde de l'IA à celui de la culture. Pour autant, **l'économie actuelle de l'IA se caractérise surtout par une concurrence extrêmement forte qui incite plus à la prédation qu'à la coopération et qu'au respect des règles.**

Dès lors, **il est nécessaire que le respect du droit se traduise** non seulement par la sécurité juridique, mais également **par un réel avantage compétitif**. Il apparaît clairement que des bases de données de qualité, complètes, expurgées de contenus douteux, nourries d'œuvres de qualité, précisément étiquetées, voire disposant d'un catalogue rare pas encore accessible, constitueront un avantage décisif pour les fournisseurs qui y auront accès. Cet atout serait d'autant plus décisif que les œuvres auraient été utilisées non pas dans la phase d'entraînement, mais dans celles d'affinage et d'ancrage.

De ce point de vue, une IA « éthique » mise sur le marché se distinguerait très avantageusement de ses concurrents, aussi bien en termes d'image de marque que de productions générées. Les entreprises européennes auraient tout à gagner à ce mouvement général, par opposition à d'autres entités qui adopteraient un comportement moins respectueux.

Recommandation n° 6 : Créer les conditions d'un réel avantage comparatif pour les fournisseurs d'IA vertueux qui sauront nouer les meilleurs accords avec les ayants droit culturels.

7. Septième principe : la diversité culturelle et la créativité humaine doivent continuer à être encouragées

Des industries culturelles enfin respectées et justement associées au succès de ce nouveau modèle économique devraient largement bénéficier des revenus générés par les IA, même si la question d'une éventuelle substitution demeure posée.

La mission d'information estime crucial de continuer à créer les conditions propices à la diversité de la création culturelle, face à la tendance intrinsèque de l'IA à l'uniformisation.

Il serait par exemple regrettable que les sources d'information d'un modèle d'IA très populaire soient limitées à un seul titre de presse, faute pour les autres d'avoir trouvé les moyens d'obtenir une rémunération. Le pluralisme, élément essentiel de notre démocratie, en serait la première victime. Il serait tout autant dommageable que la diversité musicale souffre de la sur-représentation de certains genres, ce qui nuirait à la créativité de l'ensemble de la filière musicale.

Pour la mission d'information, les revenus générés par l'IA devaient donc en partie être fléchés vers des mécanismes de promotion de la diversité de la création. De ce point de vue, le modèle des taxes affectées au CNC et des obligations de production dans le secteur audiovisuel sont des exemples éprouvés et réussis d'association de l'ensemble d'un secteur au succès des œuvres.

<p>Recommandation n° 7 : Tirer profit des revenus générés par le marché de l'IA pour promouvoir la diversité de la création culturelle et le pluralisme de la presse.</p>
--

8. Huitième principe : les créations générées par l'IA doivent être étiquetées

Face à la quantité de contenus produits par l'IA, le cadre juridique paraît à ce jour peu adapté et largement dépendant de décisions prises progressivement par les juridictions et les instances de régulation. La position constante des personnes entendues durant la mission d'information est que **les contenus générés par l'IA ne devraient pas pouvoir bénéficier du régime de la propriété intellectuelle**, ce qui est le cas actuellement.

Ce principe ne peut cependant trouver à s'appliquer que dans la mesure où **ces contenus peuvent être clairement identifiés**, que ce soit, comme il serait souhaitable, par un **marquage numérique** apposé par les acteurs de l'IA, soit par des **outils techniques**, comme celui mis en place en juin 2025 par la société Deezer.

Il s'agit, en outre, d'un enjeu essentiel pour l'information et la sensibilisation des publics.

Recommandation n° 8 : Travailler à la mise en place d'un système technique permettant d'identifier les contenus intégralement générés par l'IA.

C. DONNER SA CHANCE AU RIA ET GARANTIR SON EFFECTIVITÉ

1. Une mise en œuvre entravée

Si ses dispositions sont loin de faire l'unanimité, le RIA constitue pour l'instant le cadre dans lequel va se déployer la régulation de l'IA en Europe.

Cette législation, la seule actuellement en construction dans le monde, peut servir de modèle de régulation pour les autres pays, comme a pu l'être le RGPD. **L'Europe définirait donc pour les années à venir des standards auxquels devront se soumettre les entreprises désireuses d'accéder au marché commun.**

Dans ce contexte, il est essentiel que les outils mis en place démontrent à court terme leur efficacité, et que l'Europe ne cède pas face aux exigences américaines, telles qu'explicitées dans le mémorandum de la société OpenAI du 13 mars 2025.

Comme l'a montré la partie III du présent rapport, **la mise en œuvre du RIA pose cependant de sérieuses difficultés** compte tenu des divergences profondes entre les fournisseurs d'IA et les ayants droit culturels, mais également face au manque d'unité et de solutions partagées entre ces derniers.

Cette apparente désunion du secteur culturel ne doit pour autant pas masquer les points de consensus qui existent entre ses différentes filières. La gestion collective obligatoire et ses dérivés, comme la licence légale et l'exception compensée, sont ainsi massivement rejetées, comme l'ont montré aussi bien les auditions menées par la mission d'information que la consultation réalisée en mai 2024 par la professeure Alexandra Bensamoun, dans le cadre de la mission que lui a confiée le CSLPA.

D'une même voix, les ayants droit culturels appellent à préserver le droit de propriété exclusif du créateur sur son œuvre, qui est au fondement du droit d'auteur.

2. Une réponse graduée pour parvenir enfin à une rémunération appropriée des ayants droit culturels

L'économie de l'IA en est encore à ses débuts et l'Europe s'est lancée, avec le RIA, dans une première tentative de régulation au niveau mondial qui

se heurte tant à un secteur émergent et très concurrentiel qu'à des logiques étatiques de puissance.

Dans ce contexte, la mission d'information estime qu'il est nécessaire de laisser encore un peu de temps à la concertation, avant de chercher des moyens juridiques supplémentaires pour concilier économie de l'IA et droit d'auteur.

Elle propose en conséquence une méthode, sous la forme d'une réponse graduée en trois temps.

a) Premier temps : attendre les résultats du cycle de concertation entre les développeurs d'IA et les ayants droit culturels

Le 23 avril 2025, à l'initiative conjointe de la ministre de la culture et de la ministre déléguée chargée de l'IA et du numérique, un cycle de concertation a été lancé entre quinze représentants des développeurs de modèles d'IA générative et dix-sept représentants d'ayants droit des filières de la culture et des médias.

D'ici le mois de novembre, cette concertation devrait aborder quatre grandes thématiques :

- Identifier les besoins en données culturelles afin d'en faciliter la valorisation ;
- Forces et faiblesses des différentes modalités de rémunération et de contractualisation des données culturelles pour l'IA ;
- Retours d'expérience sur la conclusion d'accords dans un environnement concurrentiel ;
- Défis et perspectives pour un exercice efficace de l'opt-out.

L'objectif ambitieux affiché est de parvenir à une forme de consensus sur la valorisation des données et les modalités de rémunération, tant juridiquement qu'économiquement. Au vu des auditions déjà menées par la mission d'information, **la tâche paraît complexe**, tant des positions très divergentes ont été exposées.

Cependant, cette concertation a le mérite de rassembler en un même lieu des parties prenantes entre lesquels le contact n'a jamais pu être clairement établi. Il n'est donc pas interdit d'espérer qu'un accord puisse être finalement trouvé, à la satisfaction de tous.

Pour la mission d'information, la réussite de cette concertation, dont les conclusions sont attendues à l'automne prochain, serait le meilleur des scénarios.

b) Deuxième temps : en cas d'échec de la concertation, inscrire dans la loi une présomption d'utilisation des données

En cas d'échec de la concertation, et face à la difficulté de faire évoluer rapidement une législation européenne qui viendra tout juste d'entrer en application, **la mission d'information envisage de prendre une initiative législative destinée à donner toute son effectivité au RIA.**

La professeure Alexandra Bensamoun, dans son rapport précité pour le CSLPA, a proposé une analyse très aboutie des outils juridiques qui pourraient être mis en place pour parvenir à cet objectif.

La difficulté actuelle réside dans l'impossibilité pour les ayants droit culturels de s'assurer que les fournisseurs d'IA ont ou non utilisé leurs données. La partie IV du présent rapport a présenté les arguments juridiques et techniques avancés pour justifier cette absence de transparence.

L'impossibilité de prouver l'usage conduit à **une quasi-impossibilité, pour les ayants droit, de faire valoir leurs droits.** C'est pourquoi la professeure Alexandra Bensamoun propose d'**insérer dans le code de la propriété intellectuelle une notion de présomption d'utilisation.**

Concrètement, les ayants droit pourraient **se prévaloir d'un faisceau d'indices, qui irait au-delà de la simple ressemblance entre les contenus protégés et les contenus générés**, pour présumer que leurs productions ont été utilisées à une étape ou une autre par le fournisseur d'IA. Il est essentiel de ne pas limiter les cas de présomption à la simple ressemblance, tant les modalités de traitement des données par les fournisseurs peuvent être diverses.

Le fournisseur aurait alors le choix entre **accepter cette présomption**, et donc reconnaître qu'il a été amené à utiliser ces contenus, **ou bien apporter la preuve inverse.** Les conséquences seraient en fait proches de celles obtenues par une transparence totale des contenus utilisés, rejetée par les fournisseurs, mais en préservant le caractère secret des données, puisque pourraient être créées les conditions d'une réelle confidentialité.

L'introduction de cette notion dans notre droit, qui revient à une **inversion de la charge de la preuve**, semble être compatible avec le droit européen, en application du principe d'autonomie procédurale¹ et rassemble **un très large consensus de principe des ayants droit**, qui y voient une manière de retrouver, par le biais d'une législation nationale, la transparence que le RIA ne leur semble pas en mesure de faire respecter.

Par ailleurs, la présomption pourrait constituer un puissant levier pour pousser les fournisseurs à lancer enfin des négociations sérieuses et de bonne foi avec les ayants droit.

¹ Principe introduit par la CJCE par son arrêt du 16 décembre 1976 « Rewe-Zentralfinanz ».

La mission d'information se réserve donc le droit, si les concertations menées actuellement entre les parties prenantes ne parvenaient pas à des conclusions satisfaisantes, de déposer une proposition de loi allant dans ce sens.

c) Troisième temps : en cas d'échec de la mise en œuvre d'une présomption d'utilisation, créer une taxation du chiffre d'affaires des acteurs de l'IA

Si pour des raisons diverses cette deuxième option venait à échouer, **la mission d'information ne voit pas d'autre solution que d'emprunter la voie d'un prélèvement global sur le chiffre d'affaires réalisé sur le territoire national par les différents acteurs de l'IA, fournisseurs comme déployeurs.**

Les formes de ce prélèvement pourraient être diverses : taxation et affectation à un fonds dédié, obligations d'investissement dans les secteurs impactés, etc. Les industries culturelles et créatives pourraient donc bénéficier *in fine*, mais de manière dégradée, des revenus générés en partie par leurs productions.

Dans l'esprit de la mission d'information, il s'agit d'une solution de dernier recours, qui ne serait pas respectueuse du droit d'auteur. C'est pourquoi elle envisage de fixer le taux de cette taxation à un niveau dissuasif, afin de pousser les parties prenantes à un accord.

Recommandation n° 9 : Garantir l'effectivité du droit d'auteur en suivant une réponse graduée :

- attente des conclusions, à l'automne prochain, de la concertation lancée par le ministère de la culture et le ministère de l'économie entre les fournisseurs d'IA et les ayants droit culturels ;

- en cas d'échec de cette concertation à trouver des solutions adaptées, dépôt d'une proposition de loi d'initiative sénatoriale visant à mettre en œuvre une présomption d'utilisation des contenus culturels par les fournisseurs d'IA ;

- en cas de nouvel échec, mise en place d'une taxation du chiffre d'affaires réalisé en France par les fournisseurs et déployeurs d'IA, afin de compenser le secteur culturel.

EXAMEN EN COMMISSION

MERCREDI 9 JUILLET 2025

Le compte rendu de cette réunion sera publié ultérieurement.

LISTE DES PERSONNES ENTENDUES ET DES CONTRIBUTIONS ÉCRITES

Auditions des rapporteurs

VENDREDI 21 MARS 2025

- *Société des auteurs et compositeurs dramatiques (SACD)* : **MM. Pascal ROGARD**, directeur général et **Guillaume PRIEUR**, directeur des affaires institutionnelles et européennes.

- *Centre national du cinéma et de l'image animée (CNC)* : **Mme Pauline AUGRAIN**, directrice du numérique.

VENDREDI 4 AVRIL 2025

Table ronde « Presse »

- *Alliance de la presse d'information générale (APIG)* : **M. Pierre LOUETTE**, président et **Mme Léa BOCCARA**, responsable du pôle juridique.

- *Syndicat des éditeurs de la presse magazine (SEPM)* : **Mme Julie LORIMY**, directrice générale.

- *Fédération nationale de la presse d'information spécialisée (FNPS)* : **MM. Jean-Christophe RAVEAU**, président, éditeur de PYC Media, **M. Laurent BÉRARD-QUÉLIN**, vice-président, membre du Conseil supérieur de la propriété littéraire et artistique (CSPLA) et éditeur de la Société générale de presse (SGP) et **Mme Catherine CHAGNIOT**, directrice générale.

- *Syndicat de la presse indépendante d'information en ligne (SPIIL)* : **Mmes Cécile DUBOIS**, coprésidente et **Rebecca MOREAU**, chargée de mission affaires publiques.

Table ronde « Musique »

- *Syndicat national de l'édition phonographique (SNEP)* : **M. Alexandre LASCH**, directeur général.

- *Société des auteurs, compositeurs et éditeurs de musique (SACEM)* : **MM. David EL SAYEGH**, directeur général adjoint et **Blaise MISTLER**, directeur des relations institutionnelles.

- *Ekhoscènes* : **Mme Aurélie HANNEDOUCHE**, directrice générale.

- *Syndicat des musiques actuelles (SMA)* : **Mme Pauline AUBERGER**, directrice des affaires juridiques.

LUNDI 28 AVRIL 2025

- *Association Les Voix* : **MM. Patrick KUBAN**, ancien président et fondateur et **Jean VANDECASTEELE**, adhérent, comédien artiste interprète.

Table ronde « écrivains et traducteurs »

- *Conseil permanent des écrivains (CPE)* : **Mme Maïa BENSIMON**, secrétaire.

- *Association des traducteurs adaptateurs de l'audiovisuel (ATAA)* : **Mme Stéphanie LENOIR**, présidente.

- *Association des traducteurs littéraires de France (ATLF)* : **M. Jean-François CORNU**, représentant ATLF et ATAA.

- *Association pour la promotion de la traduction littéraire (ATLAS)* : **Mme Margot NGUYEN BÉRAUD**, présidente.

- *Mac Guff* : **MM. Rodolphe CHABRIER**, président directeur général et **Philippe SONRIER**, directeur général.

- *Université Paris Saclay* : **Mme Alexandra BENSAMOUN**, professeure de droit privé, personnalité qualifiée au Conseil supérieur de la propriété littéraire et artistique (CSPLA).

- *Autorité de régulation de la communication audiovisuelle et numérique (ARCOM)* : **Mme Pauline COMBREDET-BLASSEL**, directrice générale adjointe, **MM. Antoine BOILLEY** et **M. Benoît LOUTREL**, membres de l'Arcom.

Table ronde « auteurs et interprètes »

- *Société civile des auteurs multimédia (SCAM)* : **MM. Nicolas MAZARS**, directeur des affaires juridiques et institutionnelles et **Vianney BAUDEU**, conseiller affaires institutionnelles et européennes.

- *Société civile pour l'administration des droits des artistes et musiciens interprètes (ADAMI)* : **Mme Elizabeth LE HOT**, directrice générale - gérante.

- *Société de perception et de distribution des droits des artistes-interprètes (SPEDIDAM)* : **MM. Guillaume DAMERVAL**, directeur général - gérant et **Benoît GALOPIN**, directeur des affaires juridiques.

- *Syndicat national de l'édition (SNE)* : **MM. Vincent MONTAGNE**, président et **Renaud LEFEBVRE**, directeur général.

- *Autorité de la concurrence* : **MM. Thibaud VERGÉ**, vice-président et **Yann GUTHMANN**, chef du service de l'économie numérique.

MARDI 29 AVRIL 2025

Parlement européen : **Mme Laurence FARRENG**, députée européenne.

LUNDI 5 MAI 2025

Table ronde « réalisateurs et scénaristes »

- *Société des réalisateurs de films (SRF)* : **Mmes Marine FRANÇEN**, coprésidente, **Julie FABIANI**, déléguée générale adjointe et **Axelle ROPERT**, secrétaire, cinéaste.

- *Société civile des auteurs-réalisateurs-producteurs (ARP)* : **Mme Nathalie MARCHAK**, vice-présidente, **M. Radu MIHAILEANU**, vice-président, cinéaste, **M. Mathieu RIPKA**, délégué général et **Mme Joyce DARDANNE**, déléguée générale adjointe.

- *Syndicat des scénaristes* : **MM. Romain PROTAT**, co-fondateur et **Alain MOREAU**, membre.

Table ronde « arts visuels »

- *CIPAC - Fédération des professionnels de l'art contemporain* : **M. Ludovic JULIÉ**, secrétaire général.

- *Fédération des réseaux et associations d'artistes plasticiennes et plasticiens (FRAAP)* : **Mmes Paméla DORIVAL**, présidente et **Élodie LOMBARDE**, déléguée générale.

- *Société des auteurs dans les arts graphiques et plastiques (ADAGP)* : **M. Thierry MAILLARD**, directeur juridique.

- *Société des auteurs des arts visuels et de l'image fixe (SAIF)* : **M. Olivier BRILLANCEAU**, directeur général – gérant.

- *La Charte des illustrateur·rices ludiques* : **Mmes Maud CHALMEL**, membre et illustratrice et **Stéphanie LE CAM**, maître de conférences en droit privé.

- *Collectif Obvious* : **M. Hugo CASELLES-DUPRÉ**, docteur en machine learning.

VENDREDI 23 MAI 2025

- *Mistral AI* : **Mme Audrey HERBLIN-STOOP**, directrice des affaires publiques et **M. Cyriaque DUBOIS**, chargé d'affaires publiques.

- *Deezer* : **MM. Ludovic POUILLY**, vice-président principal des relations avec l'industrie musicale et les institutions et **Aurélien HÉRAULT**, directeur de l'innovation.

- *Google France* : **Mmes Sarah CLEDY**, responsable des relations institutionnelles et **Sarah BOITEUX**, responsable des affaires gouvernementales et politiques publiques, et **M. Thibault GUIROY**, responsable relations institutionnelles et politiques publiques.

LUNDI 2 JUIN 2025

- *Aday* : **M. Jean-Frédéric FARNY**, directeur général.

- *Assurance formation des activités du spectacle (AFDAS) et Audiens* : **MM. Jack AUBERT**, directeur général adjoint à l'AFDAS, **Jean CONDÉ**, directeur de l'observation, de la prospective de l'emploi et de la certification à l'AFDAS, **Frédéric OLIVENNES**, directeur général chez Audiens et **Philippe DEGARDIN**, responsable datalab chez Audiens.

- **Mme Mélanie CLÉMENT-FONTAINE**, professeure en droit privé à la faculté de droit et science politique de l'Université de Versailles Saint-Quentin-en-Yvelines (UVSQ) et directrice du laboratoire Droit des affaires et des nouvelles technologies (DANTE).

- *Spotify* : **MM. Antoine MONIN**, directeur général France et Benelux, **Romain TAKEO BOUYER**, global head of content analytics et **Mme Olivia REGNIER**, directrice des affaires européennes.

- *Union des photographes professionnels (UPP)* : **M. Matthieu BAUDEAU**, président.

LUNDI 16 JUIN 2025

Table ronde « producteurs de cinéma et de télévision »

- *Procirep & Angoa* : **M. Idzard VAN DER PUYL**, directeur général.

- *Eurocinema* : **Mme Juliette PRISSARD**, déléguée générale.

- *Association des producteurs indépendants (API)* : **Mme Hortense DE LABRIFFE**, déléguée générale.

- *Syndicats des producteurs indépendants (SPI)* : **Mme Emmanuelle MAUGER**, déléguée générale adjointe.

- *Union syndicale de la production audiovisuelle (USPA)* : **Mme Catherine LEBAILLY**, déléguée aux affaires juridiques.

- *Université Paris-I Panthéon-Sorbonne* : **Mme Joëlle FARCHY**, professeure des universités spécialiste des industries culturelles et du numérique.

- *Ministère de la Culture – Direction générale des médias et des industries culturelles (DGMIC)* : **Mme Florence PHILBERT**, directrice générale et **M. Jean-Gabriel MINEL**, délégué aux entreprises culturelles par intérim et chargé de mission auprès de la directrice générale pour les industries culturelles et créatives et l'intelligence artificielle.

Audition plénière

MERCREDI 7 MAI 2025

Mme Alexandra BENSAMOUN, professeure de droit privé à l'Université Paris-Saclay, personnalité qualifiée au Conseil supérieur de la propriété littéraire et artistique (CSPLA).

CONTRIBUTION ÉCRITE

➤ *Open AI*

DÉPLACEMENT

LUNDI 12 MAI 2025 – BRUXELLES (BELGIQUE)

- *Représentation permanente de la France auprès de l'Union européenne* :
M. Philippe LÉGLISE-COSTA, ambassadeur, représentant permanent.

- *Parlement européen* : **M. Axel VOSS**, député européen.

- *Commission européenne - DG Connect* : **Mmes Emmanuelle DU CHALARD**, cheffe d'unité « droit d'auteur », **Yordanka IVANOVA**, experte en droit d'auteur et **M. Kilian GROSS**, chef d'unité, A2 - règlement et conformité en matière d'intelligence artificielle.

- *Commission européenne - Direction générale de l'éducation, de la jeunesse, du sport et de la culture* : **M. Georg HAEUSLER**, directeur.

TABLEAU DE MISE EN ŒUVRE ET DE SUIVI DES RECOMMANDATIONS

N°	Recommandations	Acteurs concernés	Calendrier prévisionnel	Support
1	Réaffirmer et garantir le droit à rémunération des ayants droit culturels pour l'utilisation de leurs contenus par les fournisseurs d'IA.	Commission européenne et Gouvernement	2025-2026	Application du RIA par code de bonne pratique ou actes exécutoires de la Commission européenne
2	Garantir la transparence complète des données utilisées par les fournisseurs d'IA.	Commission européenne et Gouvernement	2025-2026	Application du RIA
3	Définir des modalités de rémunération qui soient fonction des flux de revenus générés par les fournisseurs et déployeurs d'IA.	Gouvernement, secteurs culturels et entreprises de la tech	2026	Accord ou cadre européen
4	Inciter le secteur culturel et celui de la presse à constituer des bases de données larges et de qualité, facilement exploitables par les fournisseurs, assorties de conditions d'utilisation précisément définies.	Secteurs culturels et entreprises de la tech	2026	Accord
5	Parvenir à un règlement financier pour les usages passés des contenus culturels, afin de compenser les ayants droit culturels et sécuriser juridiquement les fournisseurs d'IA.	Secteurs culturels et entreprises de la tech	2026	Accord
6	Créer les conditions d'un réel avantage comparatif pour les fournisseurs d'IA vertueux qui sauront nouer les meilleurs accords avec les ayants droit culturels.	Secteurs culturels et entreprises de la tech	2026	Accord
7	Tirer profit des revenus générés par le marché de l'IA pour promouvoir la diversité de la création culturelle et le pluralisme de la presse.	Secteurs culturels et Gouvernement	2026	Accord ou loi
8	Travailler à la mise en place d'un système technique permettant d'identifier les contenus intégralement générés par l'IA.	Commission européenne	2026	Règlement européen

N°	Recommandations	Acteurs concernés	Calendrier prévisionnel	Support
9	<p>Garantir l'effectivité du droit d'auteur en suivant une réponse graduée :</p> <ul style="list-style-type: none">- attente des conclusions, à l'automne prochain, de la concertation lancée par le ministère de la culture et le ministère de l'économie entre les fournisseurs d'IA et les ayants droit culturels ;- en cas d'échec de cette concertation à trouver des solutions adaptées, dépôt d'une proposition de loi d'initiative sénatoriale visant à mettre en œuvre une présomption d'utilisation des contenus culturels par les fournisseurs d'IA ;- en cas de nouvel échec, mise en place d'une taxation du chiffre d'affaires réalisé en France par les fournisseurs et déployeurs d'IA, afin de compenser le secteur culturel.	Gouvernement et Parlement	2026-2027	Accord, loi ou projet de loi de finances